A structural investigation of mechanism in human Glutathione Transferase Omega 1

Joseph Brock

BMedSc(Hons) *Australian National University*

Australian National University

A thesis submitted for the degree of
Doctor of Philosophy of Chemistry of
The Australian National University

September, 2011

Declaration

This thesis is an account of research undertaken between February 2007 and September 2011 at the Research School of Chemistry, The Australian National University, Canberra, Australia.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Joseph Brock

September 2011

Acknowledgements

I would like to offer my heartfelt gratitude and thanks to everyone who has helped me over the last 4 and a half years. While there are so many, special thanks must go to my supervisors, Aaron Oakley and David Ollis. Aaron, for sticking with me until the end despite unforeseen circumstances taking us to different parts of the country, and him having so many other commitments to contend with. I would also like to express my appreciation for how much freedom I was given during this project to learn things in my own way, in my own time. Thanks for always having the confidence in me to get this done, even when I stuffed up really bad, and for always being there for me when I got really stuck. David, thanks for being so patient and understanding of my circumstances, and for always doing your outmost best to do the right thing by me. I could never have done this without both of your unconditional support and encouragement. Special thanks must also go to Paul Carr, who out of no obligation, has given hours of his time and so much patience in teaching me the intricacies of protein crystallography. I feel privileged to have had such a vast body of knowledge I could turn to for help whenever I needed it. Thanks so much.

On a practical level, thanks to everyone in both the Ollis and Otting research groups who have unhesitatingly offered me their expertise and friendship. Much of the skills I have learnt are from the informal teachings of people like Karin, Kiyoshi, Ruhu, Hiromasa, Xi, Tracy and Jian-Wei, and for this I will always be grateful. Looking back on things now, I was so lucky to have so many outstanding research scientists who were happy to drop everything to show me anything I didn't know how to do. I was also lucky to have been surrounded by so many other students who made doing research at the RSC so enjoyable. If not for people like Colin, Chris, Morgan, Silvia and Amalia, I don't think I could have stuck at this as long as I did.

I must also thank my behind the scenes supervisor, Phil Board, and everyone in his research group. Phil's expertise on GST enzymes seems endless and he was always willing to share it with me, which proved invaluable to my understanding as my research progressed. I would also like to specially thank Marjorie Coggan who put in so much patience and time into teaching me the skills of molecular biology, cloning and protein purification at

the beginning of my project. I am lucky that the standard for my work was set so high from the very beginning.

On a personal level, thankyou to all my friends and family who have stuck with me through this. To my Mum and Dad, your unconditional love and support means the world to me. I love you both immensely and could never have done this without you. Thankyou also to my big sister Mindy, for being so cool and reminding me that your life can be anything you want it to. You have always inspired me to do the best I can possibly do. For all my close friends, I can't express how much your friendship means to me. You all kick ass and I'm so grateful that you stuck with me through this, especially in these past several months when I have been too busy to hang out or do anything.

I also need to give a big shout out to everyone at Lonsdale Street Roasters! All the free coffee you gave me and your welcome of a total stranger that hogged your limited tables for hours at a time means the world to me. Making so many new friends because of this was an awesome surprise. Besides, being able to write so much of this thing at your coffee shop really helped me get out of a ditch I felt I was in, that much caffeine will do that to you I guess.

Finally to George, you know I couldn't have done this without you. I will always love you and remember everything we have gone through together with fondness. This is as much your achievement as it is mine.

On a formal note, I would like to gratefully acknowledge the Australian National University, the Research School of Chemistry and National Health and Medical Research Council for their financial assistance of this research. This dissertation was written using the typesetting program Lyx (http://www.lyx.org/). Unless otherwise noted, figures were made using the programs Pymol (http://www.pymol.org/), ChemBioDraw Ultra (http://www.cambridgesoft.com/) and Adobe Photoshop (http://www.adobe.com/au/).

Abstract

This thesis investigates the structural details of catalysis associated with the human enzyme, Glutathione Transferase Omega 1 (hGSTO1-1), a member of the most recently discovered class of Glutathione Transferase. The human genome contains two Omega class genes (*hGSTO1* and *2*) and their transcribed products have been found at varied levels within almost every tissue of the human body. The crystal structure of hGSTO1-1 displays a distinguishing cysteine residue in the active site, a feature that undoubtedly plays a role in the novel biochemistry it displays. Of most interest is the hGSTO1-1 mediated reduction of mono- and dimethylated arsenic species that result from human arsenic exposure. The enzyme is also essential for the enzymatic detoxification of alpha-haloketones to acetophenones and exhibits a novel dehydroascorbate reductase activity. The following dissertation reports major progress in understanding the details of these chemical mechanisms. By observing these reactions within the crystallised enzyme using X-ray diffraction, several states within these catalytic cycles have been observed in great detail. Subsequently, hypotheses have been drawn about the product inhibition inferred during dehydroascorbate reduction, the existence of glutathione intermediates inherent to its interaction with methylated arsenicals, and the fundamental understanding of Omega mediated catalysis obtained by observing the enzymatic complex of oxidized glutathione and S-(4-Nitrophenacyl)glutathione simultaneously. In addition, adaptation of experimental conditions has led to the discovery of two new crystal forms with which to study the structural dynamics of this protein and its chemistry. The structure of a naturally occurring polymorphism involving the deletion of glutamate 155 has also been solved, revealing pronounced structural plasticity that correlates well with the apparent instability of the mutant enzyme *in vivo*. Discussions explore the broad implications for interpreting the role of hGSTO1-1 in the pathomechanisms of human disease states such as Alzheimer's disease and the biological mechanisms of arsenics' newly discovered potential as an anti-cancer agent.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

4NPG S-(4-Nitrophenacyl)glutathione

AAO Age at onset

AEBSF 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride

APL acute promyelocytic leukemia

ArsC arsenate reductase enzyme

AtDHAR *Aribidopsis thaliana* dehydroascorbate reductase

ATO Arsenic trioxide

ATP Adenosine triphosphate

ATRA *all-trans* retinoic acid

BLAST Basic Local Alignement Search Tool

BRF Biomolecular Resource Facility

CCD Conserved Domain Database

CDNB 1-chloro-2,4-dinitrobenzene

CML Chronic myeloid leukemia

CRID Cytokine Release Inhibitory Drug

DHA dehydroascorbate

DHAR Dehydroascorbate reductase

DMA^{III} dimethylarsinite

DMA^{V} dimethylarsinate

DMSO Dimethyl sulfoxide

DNA Deoxyribonucleic acid

DTT 1,4-dithiothreitol

E.coli Escherichia coli bacteria

ESI-TOF MS electrospray ionisation time of flight mass spectrometry

FPLC Fast Protein Liquid Chromotography

Grx Glutaredoxin enzyme

GSH glutathione

GSSG oxidised glutathione

GST glutathione transferase

hGSTO2 human glutathione transferase Omega 2 gene

hAS3MT human arsenic 3+ methyl transferase enzyme

HDNB 1-hydroxy-2,4-dinitrobenzene

HED 2-hydroxyethyl disulfide

hGSTO1 human glutathione transferase Omega 1 gene

hGSTO1-1 human Glutahione Transferase Omega 1 protein (homodimer)

hGSTO2-2 human Glutathione Transferase Omega 2 protein (homodimer)

HPLC high pressure liquid chromatography

iAs inorganic arsenic

ICP-MS inductively charged plasma mass spectrometry

IL-1 Interleukin 1

IPTG Isopropyl β-D-1-thiogalactopyranoside

kDa kilo Dalton

LB luria broth

LICs Leukemia Initiating cells

LLG log-likelihood gain

LPS lipopolysaccharide

MAD multiple-wavelength anomalous dispersion

MMA^{III} monomethylarsenite

MMA^{V} monomethylarsonate

mRNA messenger Ribonucleic acid

MWCO molecular weight cut off

NADPH Nicotinamide adenine dinucleotide phosphate

NCBI National Centre for Biotechnology Information

NCS non-crystallographic symmetry

NEB New England Biosciences

Ni-NTA Nickel nitrilotriacetic acid resin

NMR nuclear magnetic resonace

pHUE Histidine-tagged Ubiquitin Expression vector

PML Promyelocytic leukemia gene

PML Promyelocytic leukemia protein

PNP Purine Nucleoside Phosphorylase

PTPases protein tyrosine phosphatases

RARα Retinoic acid receptor alpha gene

RARα Retinoic acid receptor alpha protein

rmsd root mean square deviation

RNA Ribonucleic acid

ROS Reactive oxygen species

SDAH semi-dehydroascorbate

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SET single elelctron transfer

SNP single nucleotide polymorphism

SSM secondary structure matching

SSRL Stanford Synchrotron Radiation Lightsource

SUMOs Small ubiquitin like protein modifiers

TCEP tris(2-carboxyethyl)phosphine

Usp2-cc deubiquitylating protease

WHO World Health Organisation

Chapter 1

Introduction

The biochemical transformations of detoxification were first classified as either Phase I (functionalisation) or phase II (conjugation) (Figure 1.1) forms of enzymatic catalysis over 60 years ago [1, 2]. Since then, much has been learnt about the biological macromolecules responsible for both of these transformations. Of the first, the predominant members by far are those of the cytochrome P450 family, estimated to have a hand in the metabolism of at least 60% of drugs ingested by the human body [1]. Those belonging to the second phase of enzymatic detoxification are so named due to their substrates frequently arising from the chemical alterations achieved during the first. Their mode of action is characterised by the subsequent conjugation of bulky and water soluble moieties to these compounds. This is a process of which the glutathione transferase enzymes (GSTs) are an integral part, and generally facilitate their subsequent extrusion (phase III detoxification) (Figure 1.1).

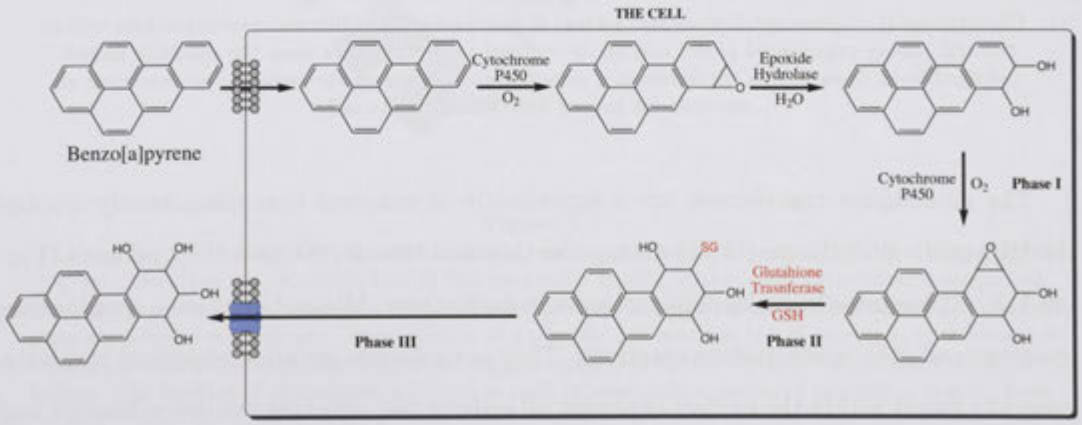


Figure 1.1
Although numerous other metabolites are also formed, shown above is one possible fate of benzo[a]pyrene upon entering a cell. Due to its hydrophobic character, the xenobiotic freely diffuses across the phospholipid membrane before becoming the subject of Cytochrome P450 and Epoxide Hydrolase Phase I catalysis. The activated hydrocarbon ring is subsequently subjected to phase II nucleophilic attack by a glutathione transferase (red). The glutathione conjugate can then be extruded by the cell using a phase III transmembrane active transporter. Such conjugates are eventually excreted from the body as mercapturic acids [3].

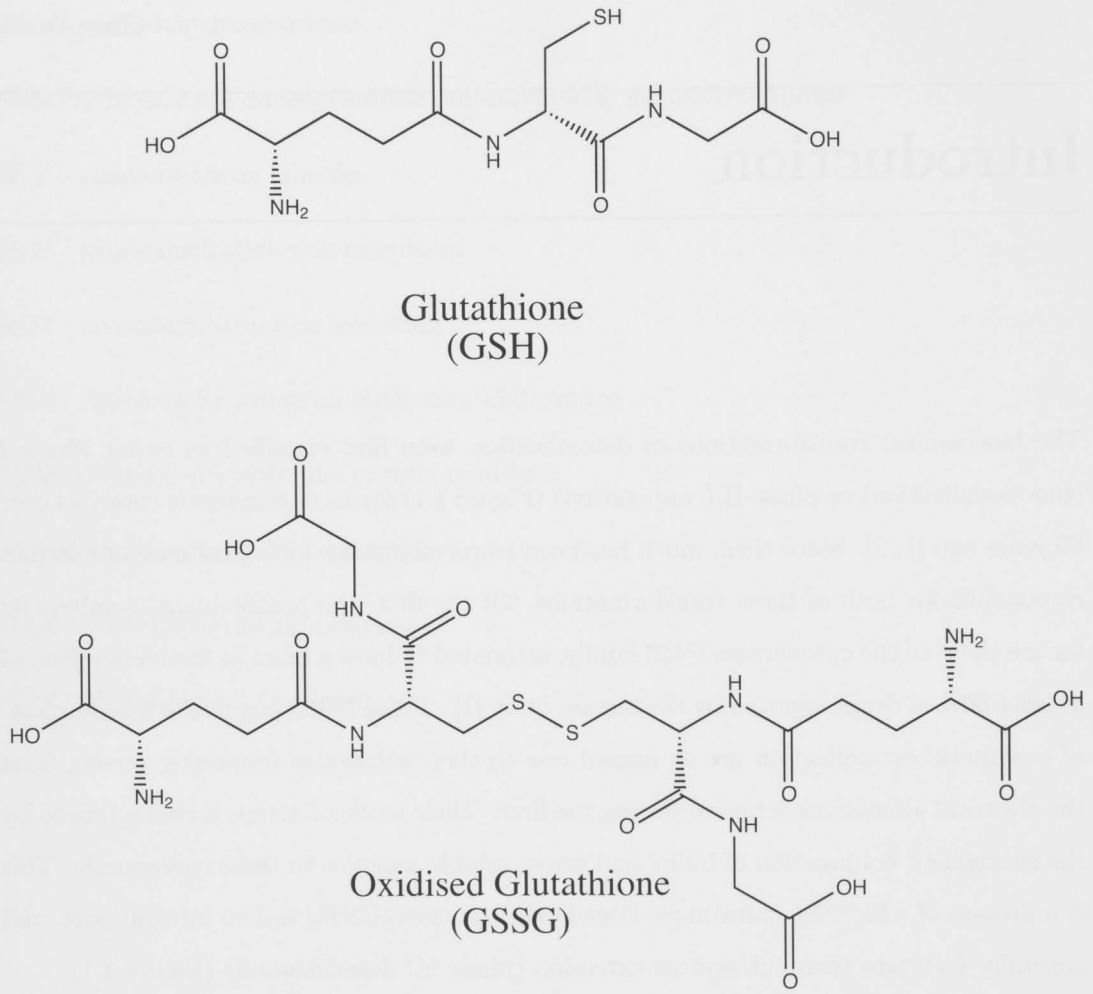


Figure 1.2
Glutathione (L-γ-glutamyl-L-cysteinylglycine) is maintained at millimolar concentrations within the cell. Being maintained principally in its reduced (upper), rather than the disulfide linked, oxidized form (lower), this tripeptide is a principal component of the reducing environment so essential for proper functioning of the cell.

The glutathione transferases are a superfamily of enzymes that ubiquitously employ the tripeptide glutathione (GSH) during the chemical transformations they perform (Figure 1.2). These include nucleophilic aromatic substitutions, Michael additions, epoxide ring openings and even isomerization reactions. They are a diverse group of biological catalysts, generally found within the cytosol of almost all cellular life, affecting the detoxification and disposition of a broad range of chemical species, including carcinogens and drugs. Their substrates are typically electrophilic compounds with internal leaving groups, however, they are also essential for the removal of reactive oxygen species and the regeneration of S-thiolated proteins, two characteristic artifacts of oxidative stress [3].

Including Omega, the cytosolic GSTs are divided into six evolutionary distinct classes

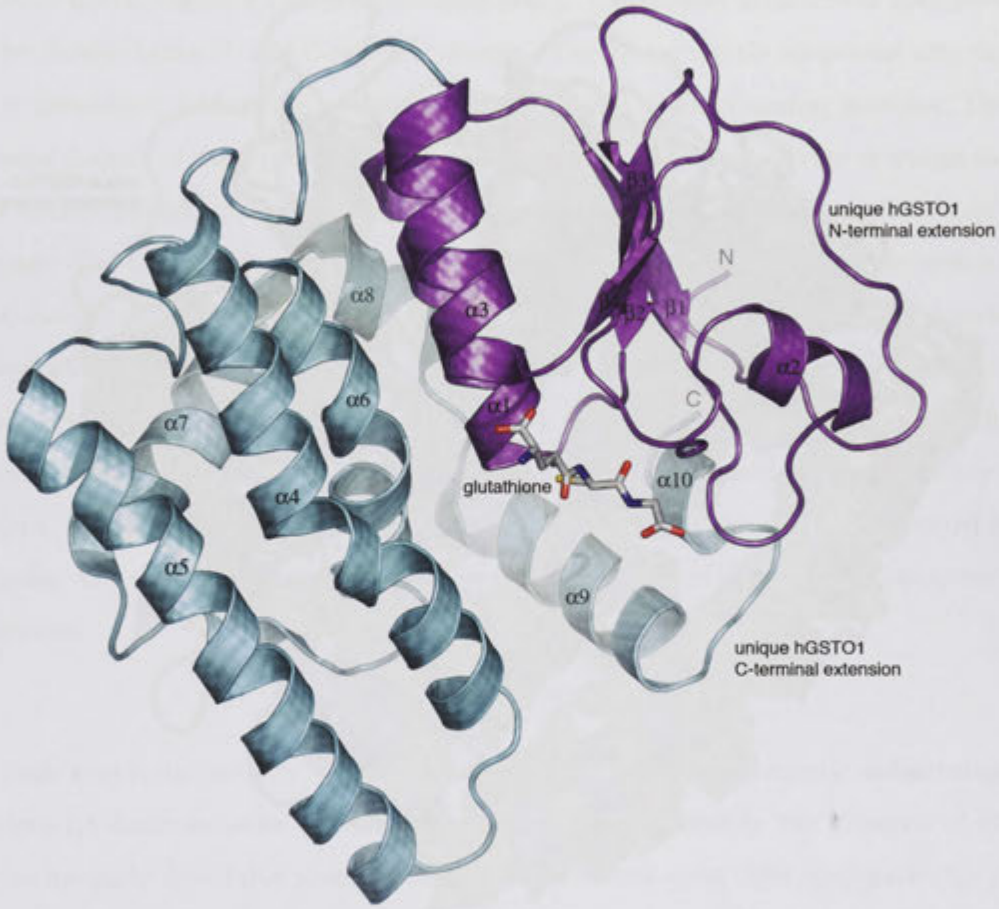


Figure 1.3

The topology of the hGSTO1 structural unit is represented with a cartoon (PDB identifier: 1EEM) [4] demonstrating the canonical fold of this enzymatic family. The N- and C-terminal domains are coloured magenta and cyan respectively. The unique extensions of these domains within the Omega class are labelled accordingly. These consist of a random coil motif at the N-terminus, that forms an unusual interaction with the C-terminal region, comprised in this instance by the novel $\alpha 9 - 10$ helices. The binding of glutathione is shown in stick representation, coloured according to atom type.

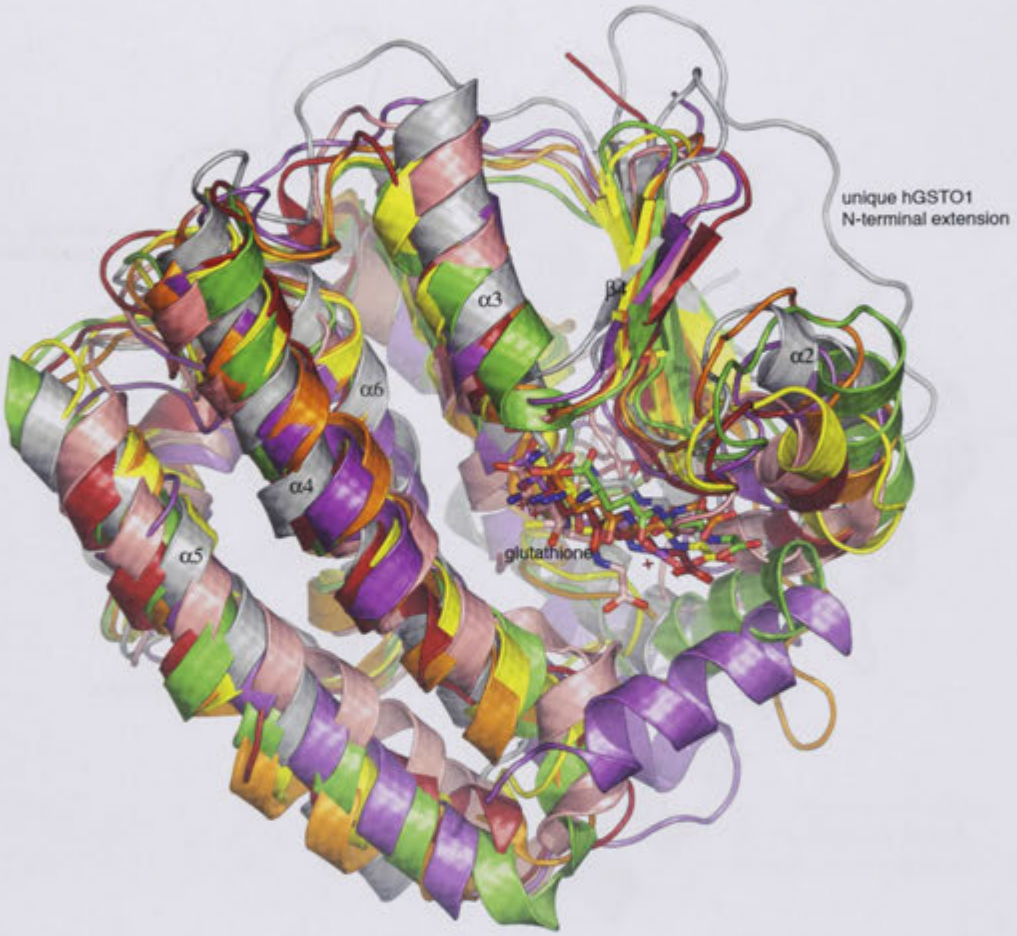


Figure 1.4
A structural comparison of representative human cytosolic GSTs with bound glutathione. Overlaid in cartoon representation are equivalent subunits of hGSTA3-3 (green, PDB identifier: 1TDI) [5], hGSTM1a-1a (orange, PDB identifier: 1XW6) [6], hGSTP1-1 (yellow, PDB identifier: 5GSS) [7], hGSTT2-2 (magenta, PDB identifier: 1LJR) [8], hGSTS1-1 (red, PDB identifier: 1PD2) [9] and hGSTZ (salmon, PDB identifier: 1FW1) [10]. The equivalent monomer of hGSTO1-1 is shown in gray (PDB identifier: 1EEM) [4], in the same orientation as Figure 1.3 on the preceding page for reference. Correspondingly coloured molecules of GSH are shown in stick representation.

within humans. Several other microsomal classes are known, as well some others found in plants and insects. Each is defined based on amino acid sequence similarity and designated with a Greek letter (α, μ, π, θ, σ, ζ and ω for Alpha, Mu, Pi, Theta, Sigma, Zeta and Omega respectively) [4, 11]. Despite pronounced differences at this level, comparing the crystal structures of each demonstrates the shape and structure that is characteristic of all cytosolic GSTs (Figure 1.4 on the preceding page). The tertiary structure of each subunit can be divided into a N- and C-terminal domain. The former clearly adopts the alternating $\beta - \alpha$ thioredoxin fold shared by several other GSH and cysteine binding proteins. The C-terminal domain of these polypeptides consists of an all α-helical motif that is unique to the GST family, providing the binding site and specificity of secondary substrates (Figure 1.3 on page 3). GST quaternary structure can be homo- or heterodimeric, involving the hybridization of different subunit types from within a given class [3]. Subunit composition is designated by Arabic numerals separated with a hyphen. The human Mu class for example, displays both possibilities. hGSTM1-1 and hGSTM1-2 indicates a homodimer composed of identical type 1 subunits, and a heterodimer, composed of a type 1 and type 2 subunit, respectively. The lower case prefix 'h', indicates that the enzyme originated from humans. As a general matter, they catalyse the conjugation of glutathione to electrophilic substrates:

$$\text{GSH} + \text{R-X} \rightarrow \text{GSR} + \text{HX}$$

Such enzymatic activity is exemplified by the nucleophilic aromatic substitution of 1-chloro-2,4-dinitrobenzene (CDNB), historically used to identify the presence of GSTs due to its easily detectable absorbance change at 340 nm upon GSH conjugation to yield l-(S-glutathionyl)-2,4-dinitrobenzene [12] (Figure 1.8 on page 12). The interaction with glutathione is typically mediated via an active site tyrosine or serine. Located in the crevice formed at the boundary of the N- and C-terminal domains, this residue facilitates catalysis by stabilising the GSH associated thiolate with a hydrogen bond. A much stronger case can be argued for this side chain hydroxyl acting as a hydrogen bond donor, as opposed to a hydrogen bond acceptor, due to the pK_a of tyrosine and serine inhibiting ionisation at neutral pH. Spectroscopic investigation of the GSH thiolate confirms a shift in pK_a from 9 to 6.7 [11]. Within the Alpha class, this thiolate is also stabilised by the positively charged side chain of an arginine residue. A conserved hydrophilic binding pocket for GSH surrounds this active site residue and is termed the 'G-site', while the binding of typically hydrophobic, secondary substrates occurs in the adjacent 'H-site' [4, 11, 3].

While either positive or negative cooperatively has yet to be observed in a wild type enzyme, disruption of a putative ion pair in the N-terminal domain of hGSTP1-1 decreases affinity for GSH, concurrently triggering a positive kinetic cooperation. The C47A or C47S mutation, located in the $\beta 2 - \alpha 2$ loop of the N-terminal domain, is hypothesised to interrupt an interaction of its thiolate form with a protonated side chain of a nearby lysine. This interaction appears to be essential for conformational stability and G-site structure. While the resulting plasticity of this region decreases the affinity for the tripeptide, it also allows structural communication between subunits, with an observed increase in the Hill coefficient from 1.01 to 1.43 [13].

1.1 Glutathione Transferase Omega

Omega is the most recently discovered class of mammalian GST. There are two Omega class GST enzymes actively expressed within the human body and they have been found at varying levels within almost every human tissue examined. The enzymes are characterised by an atypical cysteine active site residue, a feature that undoubtedly contributes to the distinguishing biochemical attributes of these enzymes. Designated hGSTO1-1 and hGSTO2-2, they possess dehydroascorbate and methyl arsonate reductase activity. The former also possesses the ability to reduce phenacylglutathiones to acetophenones and has been implicated in the post-translational activation of interleukin-1β. While its isozyme has proved difficult to purify and solubilize, hGSTO1-1 has proved more amenable to study [4, 14, 15]. The research discussed in this dissertation principally focused on discovering the chemical mechanism of catalysis with which this enzyme is involved.

1.1.1 Genetics and polymorphisms

The two functional hGSTO genes were localised to adjacent regions of chromosome 10, and designated *hGSTO1* and *hGSTO2* respectively [14]. Several polymorphisms of the former were observed within human populations. Many are within non-coding regions and their effects on splicing, transcription and subsequent mRNA stability have yet to be evaluated [16]. Several that occur within coding regions have been studied in more detail [14, 17, 18, 19]. The two of most interest are a single nucleotide polymorphism (SNP) causing an alanine to aspartate substitution at position 140, and a 3 base pair deletion at the exon/intron 4 boundary, resulting in a splicing abnormality and the subsequent deletion of glutamate 155 in the mature transcript [17].

While the A140D substitution displays no change in enzymatic activity *in vitro* [17], it has been associated with an increased risk of hepatocellular carcinoma, cholangiocarcinoma and breast cancer [20], and also implicated in the closely related pathogenesis of Alzheimer's and Parkinson's disease, vascular dementia and stroke [21, 22]. Immunostaining of the atherosclerosis plaques that lead to vascular dementia reveal a co-localization of hGSTO1-1 and Interleukin 1 alpha (IL-1α) [23], supporting the notion that hGSTO1-1 may influence cerebrovascular disease via regulation of IL-1 expression in macrophages and monocytes [23, 24] (Chapter 6 on page 131).

The Δ155 polymorphism is characterized by two haplotypes involving a glutamate or lysine at position 208 and its frequency is found to fluctuate between ethnic groups, with over 10% possessing the deletion mutation in one Chinese sample [25, 26, 14]. Recombinant expression of this polymorphism in *E. Coli* resulted in an enzyme possessing enhanced activity *in vitro* and a compromised heat stability [25, 14]. The T-47D breast cancer cell line has been shown to be hemizygous for the E155 deletion and displays decreased hGSTO1-1 activity. Furthermore, the same study found that lymphoblastoid cell lines from individuals heterozygous for the allele had only 50% of normal activity towards the specific substrate S-(4-Nitrophenacyl)glutathione (4NPG) (Chapter 5 on page 97), indicating the enzyme is severely compromised *in vivo* [27] (Appendix A on page 181).

1.1.2 Structure

While adopting the canonical GST fold, the crystal structure of hGSTO1 reveals unique extensions of both its N- and C-termini (Figure 1.3 on page 3). The proline rich N-terminal region of approximately 19 residues is observed to undergo several hydrogen bonding interactions with the two extra α-helices of the C-terminus, forming a unique element of continuous secondary structure on its surface (Figure 1.6 on page 10). Additionally, the mode of glutathione binding has two distinctive differences from almost all other GSTs.

Firstly, the tyrosine or serine residue residing at the active site of almost all other GSTs is replaced by a cysteine within the Omega class, and is observed to form a mixed disulfide bond with glutathione in the associated crystal structure [4]. The only other instance of this GST active site motif has been in the bacterial Beta class [3].

Physiologically active as a homodimer, the cleft between the subunits of hGSTO1-1 is notably larger than that of most GSTs (Figure 1.5 on page 9), such that it's binding to other polypeptides is conceivable [4, 28]. The second distinguishing difference in GSH

binding and this feature of hGSTO1-1 quaternary structure are related. Due to the large cleft at the interface, the GSH binding to each monomer does not interact with the adjacent monomer. A salt bridge between the N-terminal amino-group of the tripeptide is typically mediated by an acidic residue on helix α4 of the adjacent monomer within almost all other GSTs. Indeed, most of the interactions at the hGSTO1-1 dimer interface are non-polar in character. There are no hydrogen bonds and the only two salt bridges, the two occurrences of K114-E91 [4] (Figure 1.6 on page 10). Such hydrophobic dimer interfaces are a structural feature characteristic of GSTs, with only the sigma class displaying atypically polar interactions in this region [3].

Unlike the dimer interface however, the H-site of hGSTO1 is an atypically polar environment. Characterised most definitively by the side chains of R183 and W222, this is suggestive that the many distinguishing features of Omega may have evolved to deal with more polar substrates than the rest of its family. Observed to be orientated with its indole nitrogen pointing directly into the binding pocket, this hydrogen bond contributed by W222 belongs to the unique C-terminal extension of the hGSTO1 polypeptide (Figure 1.7 on page 11) [4].

1.1.3 Chemistry

Quite distinct from most other GSTs, the Omega class enzymes have no appreciable activity towards CDNB[1] (Figure 1.8 on page 12). *hGSTO1* and *hGSTO2* were rather discovered via a database search of more than 10^6 expressed sequence tags (ESTs). A segment of nucleotides corresponding to the Zeta class was used as a query sequence [4]. Together with the Theta class, these are the only other mammalian GSTs known to lack activity towards this substrate. Once cloned and expressed, the resulting enzymes were found to possess 'thioltransferase' activity, a mode of catalysis quite distinct from the 'glutathione conjugation' behaviour exhibited by almost all of its family members (Figure 1.8 on page 12) [4, 17]. Such chemistry is reminiscent of the glutaredoxin enzyme family, a group of enzymes that also exhibits an active site cysteine and thioredoxin like fold [3, 30]. This change in the usual GST catalytic cycle has manifested itself as several unprecedented biochemical activities within Omega, including the rejuvenation of oxidized ascorbate (Chapter 3 on page 43), the metabolism of arsenic (Chapter 4 on page 61) and the detoxification of

[1] Although an independent investigation has reported that hGSTO2-2 does show high enzymatic activity towards CDNB [29], this claim has not been supported by the thorough characterization of this enzyme being currently performed by my colleagues (unpublished data).

Figure 1.5
A cartoon representation of the hGSTO1-1 physiological dimer as reported by Board *et. al.*, [4], viewed approximately perpendicular to the twofold axis of symmetry. The active site cysteine and bound glutathione are shown in stick representation, coloured according to atom type. Each subunit of the homodimer is coloured magenta and cyan respectively.

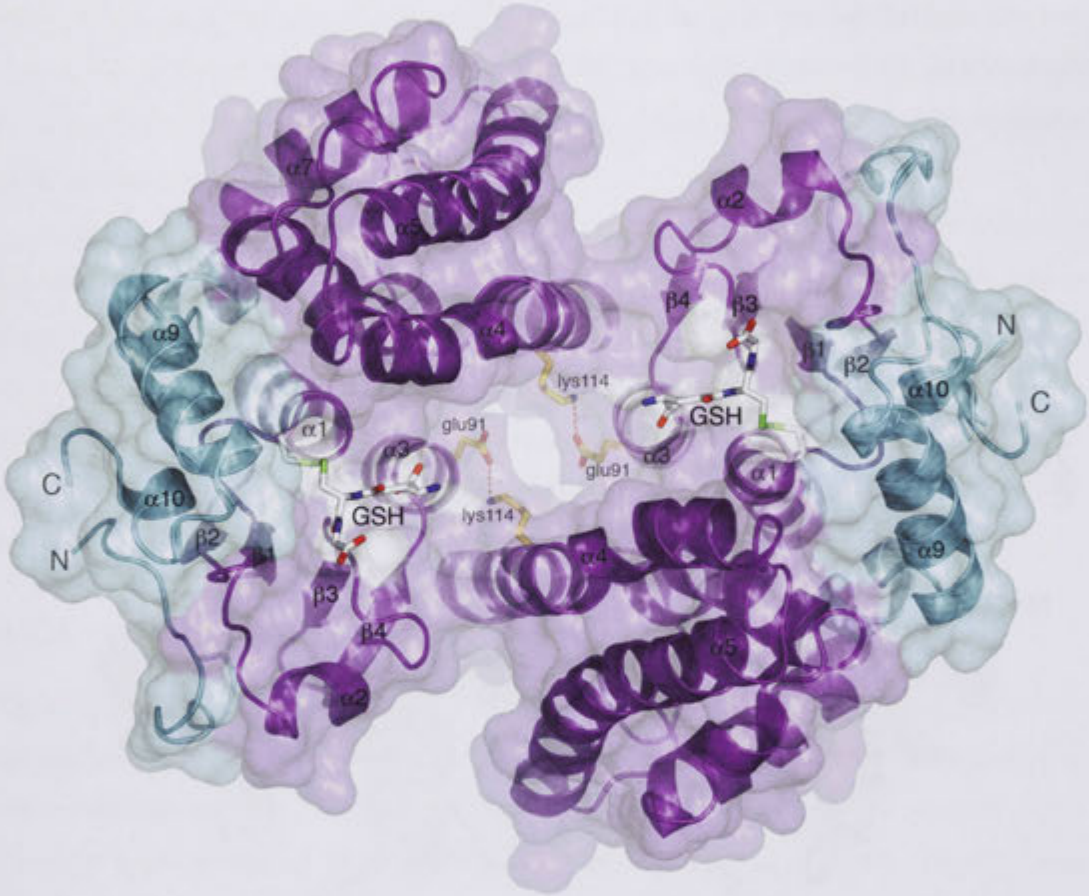
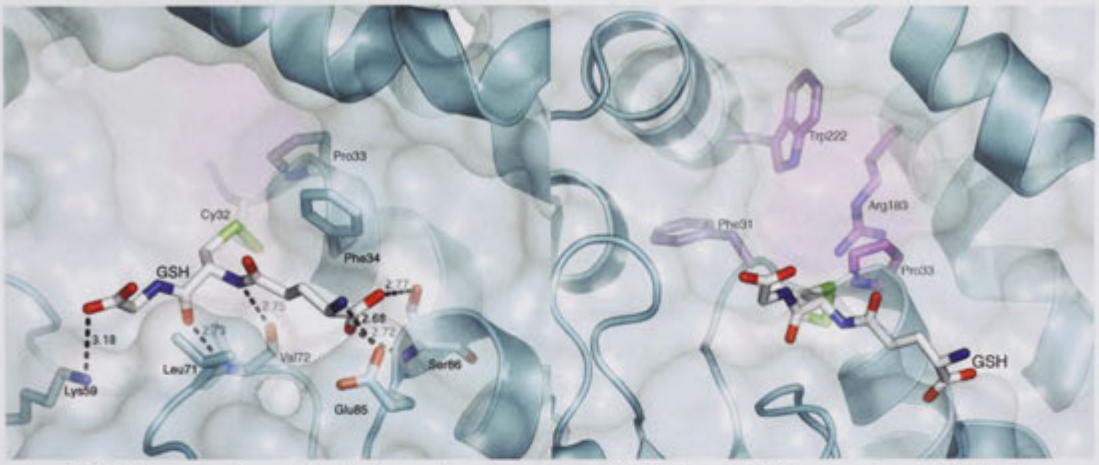


Figure 1.6

A cartoon representation of the hGSTO1-1 physiological dimer as reported by Board *et. al.*, [4], viewed approximately parallel to the twofold axis of symmetry. The structural fold common to all GSTs is displayed in magenta, with the novel N- and C-terminal extensions of hGSTO1-1 in cyan. The continuous element of secondary structure formed by their interaction is highlighted with a correspondingly coloured molecular surface. The active site cysteinyl-GSH mixed disulfide is shown in stick representation, coloured according to atom type, with the associated molecular surface coloured white. Also shown in stick representation are the lone instances of polar interactions found at the dimer interface. Highlighted in yellow, these are the two instances of the glutamate 91-lysine 114 salt bridge. The crystallographic distance of these interactions are given in Angstroms (Å).

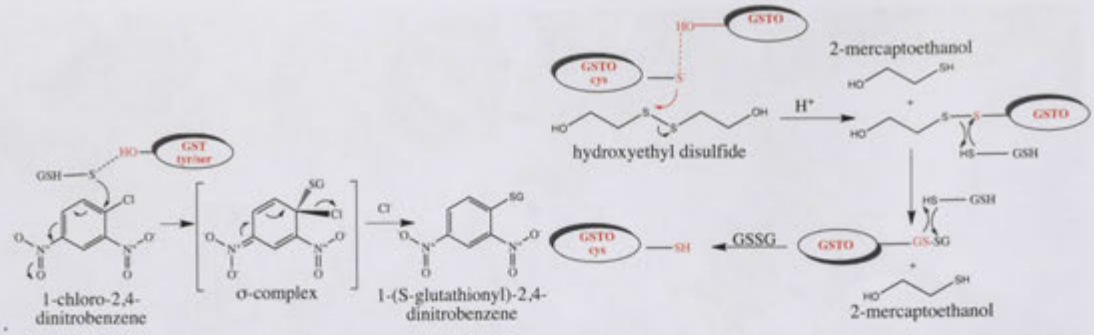


(a) The crystallographically determined 'G-site' of hGSTO1. Polar interactions with the tripeptide glutathione are indicated with a dashed black line. The crystallographic distance of each is given in Å.

(b) The well defined cavity of hGSTO1, corresponding to the 'H-site' of substrate binding in other GST classes. The relatively hydrophilic group of amino acids defining this wide and deep pocket are highlighted in magenta.

Figure 1.7
The catalytic active site of hGSTO1. The polypeptide fold is shown as a cyan coloured cartoon covered in a translucent molecular surface. Glutathione and relevant amino acids are shown in stick representation, coloured according to atom type. White and magenta are used to highlight the 'G-site', and putative 'H-site' of the enzyme, respectively [4].

phenacylglutathiones to acetophenones (Chapter 5 on page 97) [15, 17, 25]. The atypically wide orientation of monomers within the hGSTO1-1 physiological dimer suggests another application of its thioltransferase activity may be the glutathionylation or deglutathionylation of other enzymes, a mechanism commonly observed to modulate protein function [4]. Although the presence of a stabilising hydrogen bond donor has not yet been identified, this active site cysteine is crystallographically observed directly over the N-terminus of the helix-α1 dipole [4]. Facilitated in this case by the *cis*-proline residue it precedes, such positioning of a cysteine residue substantially lowers its pK_a [31]. This active site cysteine has been shown to be essential to the enzymes thioltransferase activity [17] and also the target of several covalently modifying inhibitors [32, 24] (Chapter 6 on page 131). In contrast, mutating this cysteine to alanine greatly enhances the glutathione conjugating activity of hGSTO1 with the 'universal' substrate, CDNB. Its removal presumably allows glutathione to assume its usual role as an attacking thiolate (Figure 1.8 on the following page) [17].



(a) The typical catalytic mechanism of GSTs towards the xenobiotic 1-chloro-2,4-dinitrobenzene.

(b) The atypical thioltransferase activity of the Omega class GSTs.

Figure 1.8
The catalytic mechanism of a typical GST enzyme (red) compared to that of GST Omega. Most GST classes have been discovered via their activity to the 'universal' substrate, 1-chloro-2,4-dinitrobenzene (CDNB) shown in (a) [33, 12]. Omega GSTs have little to no activity towards this xenobiotic however and preferentially display a thioltransferase activity instead (b) . This atypical mode of catalysis, a biochemical mechanism often involved in the modulation of enzymatic activity *in vivo*, is undoubtedly related to the cysteine residue characteristic of this enzymatic class. Unlike the hydroxyl moiety observed in other GST classes, the hydrogen bond donor responsible for stabilising the nucleophilic thiolate in the Omega catalytic mechanism has not as yet been identified [4].

1.1.4 Other features of GST Omega

The crystal structure of hGSTO1-1 has revealed striking similarity to a nuclear chloride channel, CLIC-1 [34, 28], and prompted a subsequent investigation as to whether the protein could form or modulate ion channels. Conducive to hGSTO1-1 being able to modulate other proteins via its thioltransferase activity, this study has revealed its ability to significantly affect the activity of Ryanodine receptor Ca^{2+} channels *in vitro*. The active site cysteine was shown to be essential to this mode of action [28].

hGSTO1-1 has also been shown to be upregulated to levels greater than tenfold in human estrogen receptor negative breast cancer cells, relative to estrogen receptor positive cells. Within breast carcinomas, their exists an inverse relationship between estrogen receptor expression and several mestatic properties, such as mobility and invasiveness [35]. Upregulation of hGSTO1-1 has also been shown to be statistically significant in cancer cells resistant to the chemotherapeutic agent *cis*-platinum [36].

1.2 Research Objectives

This project specifically aimed to elucidate the details of hGSTO1-1 mediated chemistry, the mechanism of specific inhibitors, and to gain insight as to how these interactions influ-

enced the structural dynamics of the enzyme. To achieve this, a variety of X-ray diffraction experiments were performed upon several crystal forms of the enzyme that involved various substrates, substrate analogues, inhibitors and the site directed mutagenesis of it's active site residue. This information is crucial to developing a more detailed picture of the role played by hGSTO1-1 within the body, its contribution to the disease states with which it has been associated, and how this contribution may be augmented for therapeutic purposes.

An overview of the experimental methods

Crystallographic investigation of mechanism is often hindered by the expediency of catalysis, such that the fleeting interactions of substrates are difficult to capture in the relatively long time scale required to collect a dataset. It was thus envisioned to attempt the cloning and crystallisation of some hGSTO1 active site mutations, specifically, the conversion of the active site cysteine to an alanine and a serine residue. It was also envisioned to employ the substrate analogue, glutathione sulfonic acid, in crystal growth. By thus removing the necessary reducing equivalents at various stages of the catalytic mechanism, it was hoped the enzyme could be 'stalled' in intermediary stages of catalysis that would prove more amenable to investigation by X-ray diffraction. As my research progressed, other experimental parameters were interactively altered to affect change within the crystallographically observed active site. This included adjusting the molar ratio of substrates and co-substrates relative to the enzyme, combining potential allosteric inhibitors during co-crystallisations with substrate, incubation of substrates and inhibitors with the enzyme at room temperature before crystallisation at 4°C, and the alteration of redox conditions used in purification combined with additive screens during crystallisation, eventually leading to the discovery of two new crystal forms (Chapter 6 on page 131).

This chapter is designed to give a brief introduction to the standard techniques used throughout the course of this research. More detailed information pertaining to the experimental conditions of particular experiments will be given in the relevant chapters.

2.1 Site directed mutagenesis

The alanine active site mutant of *hGSTO1* had previously been made by a colleague [17] but first needed to be transferred from the pQE30 plasmid (Qiagen), to the expression vector pHUE [37] (Figure 2.2 on page 19 and Section 2.2 on page 18), to facilitate large scale protein expression with a removable his-tag for crystallization. This was performed via a complementary restriction digest of both vectors, followed by gel visualisation, isolation and purification of the insert and vector fragments, respectively . The two DNA fragments were then ligated, before transfecting into *E. coli* (Section 4.2 on page 66). After subsequently purifying the plasmid, the active site mutation was then confirmed unequivocally by DNA sequencing (Figure 4.3 on page 67).

The cysteine to serine substitution was performed via site directed mutagenesis, using native hGSTO1 in the pGEM-T plasmid (Promega) as a template. Primers containing the desired nucleotide change were employed in a PCR reaction to perform a whole plasmid replication. As shown in the Stratagene 'QuikChange' protocol (Figure 2.1 on the facing page), this allowed the amplification of only the mutated plasmid, while template plasmid could subsequently be digested with a methylation specific restriction enzyme. After transfecting *E. coli*, the mutated plasmid was purified and the cysteine 32 to serine substitution also confirmed via DNA sequencing (Figure 3.4 on page 47), before restriction digest and ligation into the pHUE vector as above.

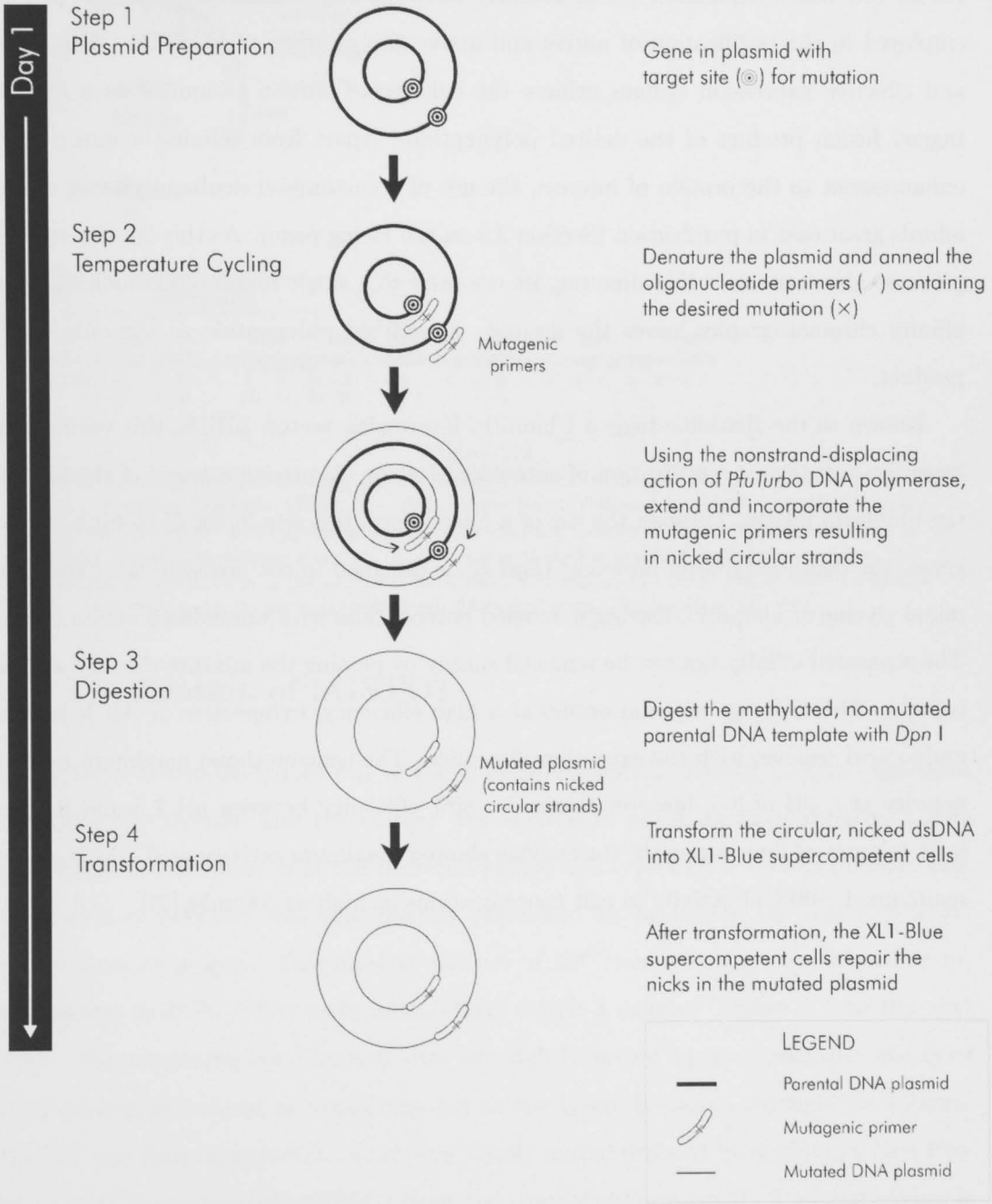


FIGURE 1 Overview of the QuikChange site-directed mutagenesis method.

Figure 2.1
The 'QuikChange' site directed mutagenesis protocol as shown in the Stratagene instruction manual (www.stratagene.com catalogue #200518).

2.2 The pHUE expression system

An *E. coli* based expression vector recently developed by Cantanzariti *et. al.*, [37] was employed in the purification of native and active site mutants of hGSTO1. This simple and effective expression system utilizes the eukaryotic protein Ubiquitin as a histidine tagged fusion product of the desired polypeptide. Apart from offering a natural yield enhancement to the protein of interest, the use of a customised deubiquitylating enzyme affords great ease in purification (Section 2.3 on the facing page). As this deubiquitylating protease also contains a Histidine tag, its use prior to a single round of immobilised metal affinity chromatography leaves the desired, unmodified polypeptide as the only soluble product.

Known as the Histidine-tagged Ubiquitin Expression vector, pHUE, this vector facilitates the quantitative purification of authentic proteins via precise removal of the histidine tag-ubiquitin fusion. Through the use of a *SacII* restriction site in the 5' cloning of insert genes, the deubiquitylating protease, Usp2-cc, is designed to cut precisely after the C terminal glycine of ubiquitin, leaving unlabeled polypeptides with unmodified amino termini. The separated affinity tag can be removed simply by passing the mixture through a Nickel column. This cleavage reaction occurs at a high efficiency, irrespective of the N-terminal amino acid residue, with the exception of proline. The enzyme shows maximum catalytic activity at a pH of 8.5, however retained >80% efficiency between pH 7.5 and 9. Similarly tolerant of ionic strength, the enzyme showed maximum activity at 0 mM NaCl, but maintained >90% of activity in salt concentrations as high as 300 mM [37].

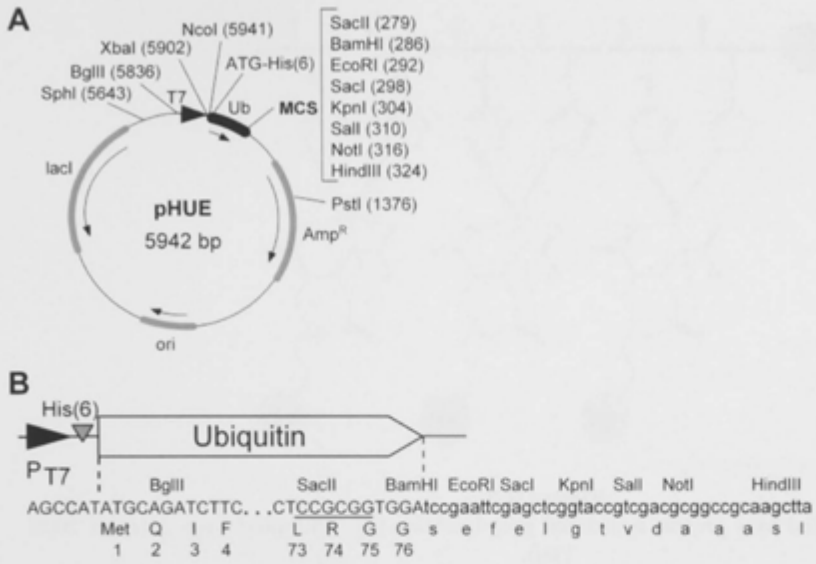


Figure 2.2

The plasmid map of Histidine-tagged Ubiquitin Expression vector, pHUE, (A). The inducible T7 RNA polymerase promoter (black triangle) is immediately followed by the His-tagged Ubiquitin (Ub) coding region. The appropriate restriction sites are indicated. Other abbreviations are as follows: Ampr, β-lactamase gene; ori, *E. coli* origin of replication; lacI, lacI repressor gene. (B) Shows the 5' and 3' ends of the ubiquitin construct. The engineered SacII site by which a precise fusion can be made is underlined. (Adapted from Catanzariti *et al.*, [37]).

2.3 Purification of hGSTO1

After growth of transfected *E. coli* to the appropriate optical density, protein expression was chemically induced and the culture was allowed to reach maximal capacity before centrifugation, resuspension of the bacterial pellets and lysis with a French pressure cell. Purification of hGSTO1 was subsequently performed by subjecting the lysate to nickel affinity chromatography. This involves the use of Ni^{2+} ions partially coordinated to nitrilotriacetic acid (Ni-NTA) resin immobilised within a column (Figure 2.3 on the next page). The remaining coordination sites are tightly bound by the imidazole nitrogens of the protein associated poly-histidine tag as the lysate is passed through the column. hGSTO1 was then separated from non-specifically bound proteins by applying a Fast Protein Liquid Chromatography (FPLC) mediated concentration gradient of soluble imidazole (Figure 2.4 on page 21). Elution of protein was identified by absorbance of aromatic amino acids at 280 nm, allowing specific fractions to be checked for the presence of hGSTO1 via SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) (Figure 2.5 on page 22). The most pure fractions were pooled and dialysed in buffer, before performing a cleavage reaction via addition of the Usp2-cc ubiquitinase (Section 2.2 on the facing page). The ubiquitin-histidine tag, ubiquitinase and any remaining impurities could then

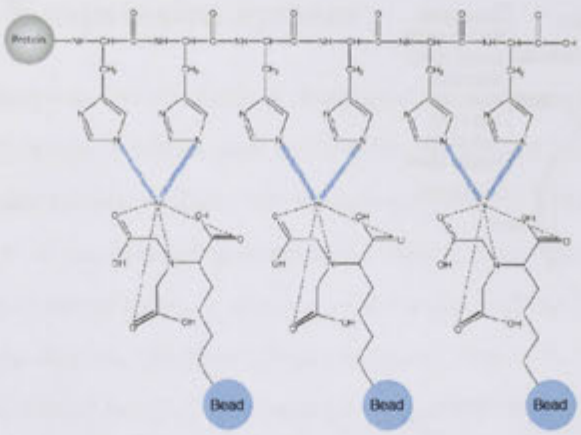


Figure 2.3
Ni-NTA resin coordinated by a protein associated poly-histidine tag. (Adapted from Astatke, 2008 [38]).

removed by another pass through the Ni-NTA column (Figure 2.5 on page 22). The resulting flow through containing only pure protein, was finally exchanged into new buffer, before concentrating and usage in crystallisation screens.

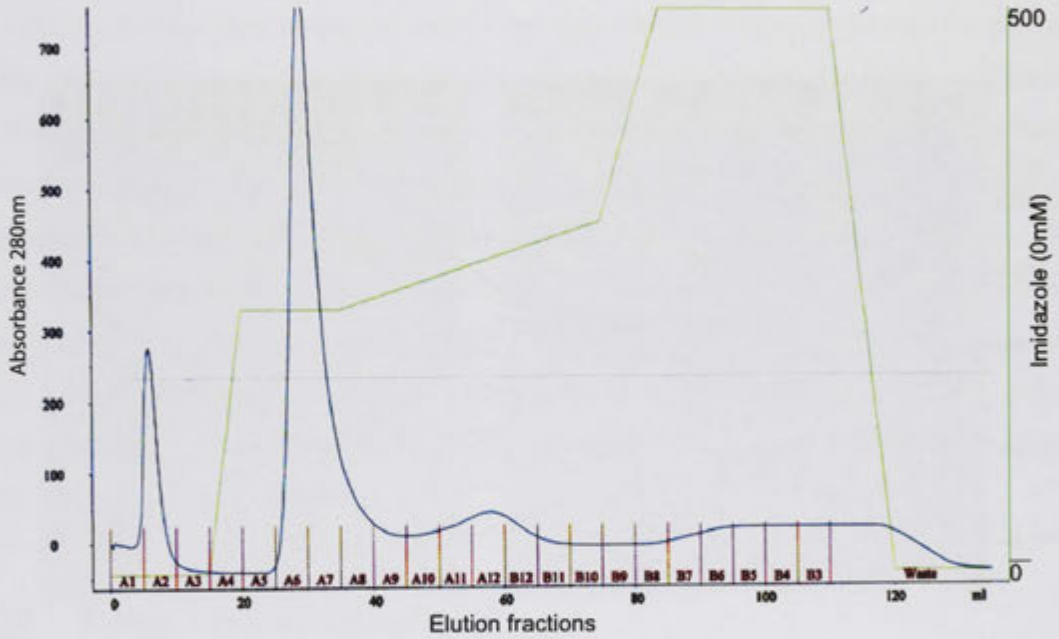


Figure 2.4

An example of the FPLC mediated purification of hGSTO1-1. The elution of protein from the 'His-trap' (GE Healthcare) Ni-NTA column is shown by absorbance of the elutant at 280 nm, indicated by the blue graph. This is shown in comparison to the imidazole concentration of the wash buffer, indicated by the green graph. The composition of each elution fraction was then anaylsed by SDS-PAGE (Figure 2.5 on the next page).

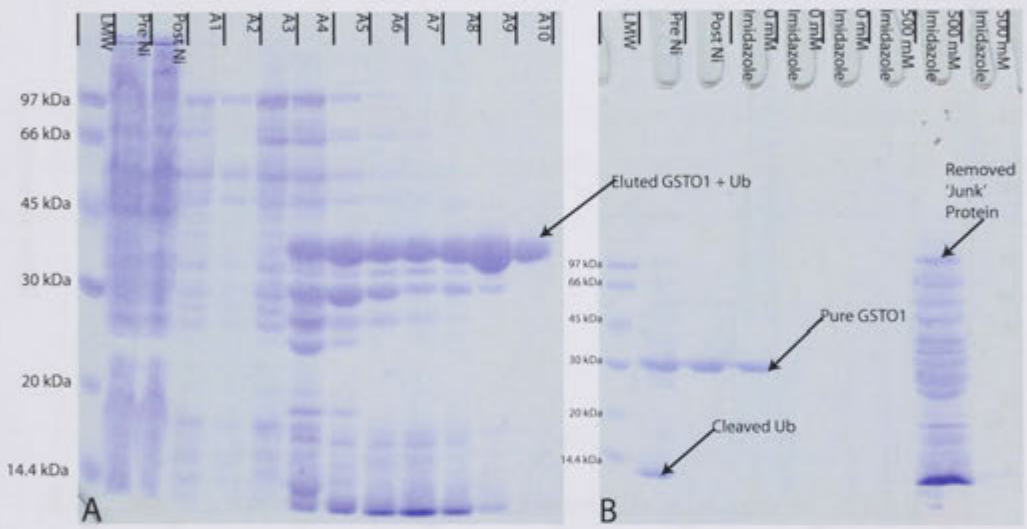


Figure 2.5

SDS-PAGE analysis of hGSTO1 purification. As imidazole is a histidine analogue, its increasing concentration corresponds to tagged protein elution. (A) Lanes A1-10 correspond to protein eluting at an imidazole concentration of ≈70 mM to ≈370 mM respectively. The hGSTO1/Ub construct is clearly observed at ≈35 kDa. Collected protein corresponding to lanes A5 and onwards were combined and digested with Usp2-cc protease. (B) Upon inspection of the column flow through (post Ni) and subsequent washes with buffer, pure hGSTO1 is observed (≈30 kDa) as the only soluble product, effectively demonstrating the efficiency of this two step process. The impurities previously present are observed to elute upon washing with excess imidazole. 2 µl of low molecular weight (LMW) protein mixture (GE Healthcare) is shown in the left most lanes of each gel. The size of its constituents are given for reference.

2.4 Purification of the Deubiquitinase, Usp2-cc

A small quantity of cloned *Usp2-cc* in pET15b (Novagen) was obtained from colleagues at the John Curtin School of Medical Research (JCSMR). 1 μL of this plasmid construct was used to transform a 50 μL aliquot of heat shock competent *E. coli* cells of strain DH5α via incubation on ice for 10 minutes, followed by heat shock in a 42°C water bath for 45 seconds[1]. The transformed cells were immediately spread onto an agar plate containing 0.2 mg/ml Ampicillin. After incubation at 37°C overnight, the resulting colonies were used to grow a starter culture of luria broth (LB) also containing 0.2 mg/ml Ampicillin. The resulting cells were spun down and used to generate a larger volume of the plasmid using a Mini Prep kit (Qiagen). 1 μL of the purified plasmid was subsequently used to transform a 50 μL aliquot of electroporation competent *E. coli* cells of strain BL21. This was then streaked out on an agar plate (0.2 mg/ml Ampicillin) and subsequently used to grow a 5L volume of transformed culture. As the pET15b-*Usp2-cc* construct also encodes an N-terminal poly histidine tag fused to the ubiquitinase, it was purified via French pressure cell lysis of cells and subsequent Ni-affinity chromatography as in Section 2.3 on page 19. Cell resuspension and post-purification dialysis however, was in a slightly altered buffer (300 mM NaCl, 50 mM Sodium Phosphate, 10 mM imidazole and 30% Glycerol, pH 8) [37].

2.5 X-ray crystallography

X-ray crystallography was the principal method used to study the structural basis of hGSTO1-1 catalysis during this investigation. It is essentially a way to visualise what happens to protein molecules and their substrates during catalysis. As the resolution of any microscopic investigation is physically limited by the wavelength of radiation used, X-rays with wavelengths of approximately 1 Å are ideal to resolve the complex bonding arrangements of atoms within crystals. Conventional forms of microscopy typically use a lens to refocus scattered light into a magnified image. The lack of a physical lens capable of refocusing X-rays however, means they must alternatively be diffracted through an ordered 3-dimensional array. This effectively amplifies the diffraction pattern, which can then be sampled in a 3-dimensional pattern of discrete spots. Application of accurate

[1]Competent cells were kindly prepared by resident lab technicians, Mr Ruhu Qui and Miss Tracy Murray.

phase information then facilitates creation of an image. The only 3-dimensional arrays for this purpose are crystals. The resulting diffraction pattern has a physical relationship to the arrangement of atoms at each lattice point of the crystal. It is a relationship that is described precisely by a mathematical operation known as the Fourier transform. Electromagnetic radiation can be described in terms of its amplitude (half the height of each wave, from peak to trough), wavelength (distance between successive peaks) and phase (lateral displacement of the wave relative to an arbitrarily chosen origin). Once this information about each X-ray beam diffracted from the crystal is known, the Fourier transform can be used like the lens of conventional light microscope, creating a picture of the protein's electron density that is repeated throughout.

The detailed theoretical proofs of X-ray crystallography and their derivations have been well established for several decades and will not be reproduced here. What follows is a brief introduction to some crystallographic methods, in the context of how they were employed in this research. For an introductory textbook to the subject, the reader is directed to '*Crystallography Made Crystal Clear*' by Gale Rhodes [39]. This intuitively written text regularly makes use of figures that visually convey the theoretical concepts involved in a simple and concise fashion. Some have been adapted below to illustrate the methods used.

2.5.1 Crystal symmetry and crystallization.

Crystals can be described as a 3-dimensional lattice of symmetrically equivalent points. Each is in an identical environment, such that an observer would have the same perspective of the crystal lattice from each. The lattice points dictate the boundaries of what can be thought of as many tiny, identical boxes, stacked tightly against each other. These boxes are known as *unit cells*, and their packing in 3-dimensions is compatible with several shapes, known as *crystal systems*. In order of descending symmetry, these are cubic ($a = b = c$ and α=β=γ=90^o), tetragonal ($a = b \neq c$ and α=β=γ=90^o), orthorhombic ($a \neq b \neq c$ and α=β=γ=90^o), hexagonal ($a = b \neq c$ and $\alpha = \beta = 90^o$, γ=120^o), rhombohedral ($a = b = c$ and $\alpha = \beta = \gamma \neq 90^o$), monoclinic ($a \neq b \neq c$ and $\alpha = \gamma = 90^o, \beta > 90^o$) and triclinic ($a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$). Several lattice centerings are also possible, known as *lattice types*. A unit cell that contains lattice points at its vertices only is known as a primitive centered lattice type. Within crystal systems other than triclinic, 3 others are also possible: body centered (an additional point in the centre), face centered (additional lattice points in the centre of each unit cell face) and a, b or c centered (two additional lattice points in

the centre of two opposing faces). The non-redundant combination of the 4 lattice types with the 7 crystal systems leads to the 14 unique types of unit cells known as the *Bravais lattices.*

The unit cell is the smallest element that can reproduce the entire crystal by purely translational symmetry, but they generally contain *internal* symmetry as well. This includes rotational and reflective symmetry elements, *i.e.,* rotation axes and mirror planes. These can also be combined with translational elements, creating screw axes and glide planes. This internal symmetry of a unit cell is described by its *space group*, a symbol that describes the combination of symmetry operations applied to the *asymmetric unit.* This is the smallest element that can reproduce the unit cell with translational *and* rotational/reflective symmetry operations. Applying a particular space group symmetry to an asymmetric unit, and then translating the resulting unit cell in 3-dimensions, generates the entire crystal lattice. In total, there are 230 different space groups compatible with 3-dimensional translational symmetry. Due to the presence of only L-configured amino acids in proteins however, this number is much lower within the crystals they form: only the 65 enantiomorphic space groups that do not contain mirror planes, glide planes or other reflective elements are relevant.

As an example, hGSTO1 readily crystallizes in the space group $P3_121$, with unit cell dimensions of 56.2 × 56.2 × 139.5 Å [4]. The capital letter 'P', designates the primitive lattice type, numbers indicate rotational symmetry, and subscript numbers indicate a screw axis. Within $P3_121$ crystals of hGSTO1, the asymmetric unit is a single protein monomer and there is 3-fold screw axis (3_1) down the unique unit cell axis. This means that an identical position within the crystal lattice is achieved by combining a 120° (360°/3) rotation, with a translation of the asymmetric unit by $^1/_3$ its length. There is also a 2-fold rotational axis of symmetry (2) perpendicular to this, happening to correspond to the 2-fold axis of the physiological dimer.

By concentrating and purifying proteins to homogeneity, before combining them with various salts, buffers and precipitants in controlled environments, they can sometimes be coaxed into forming such crystals. Since the theory of crystal formation is poorly understood, this is typically a trial and error process. Many different combinations of different chemicals are screened to find initial hits that can then be optimised. Over time, some compounds have been identified as being particularly effective. The technique of vapour diffusion is employed in this process, whereby protein is mixed with various

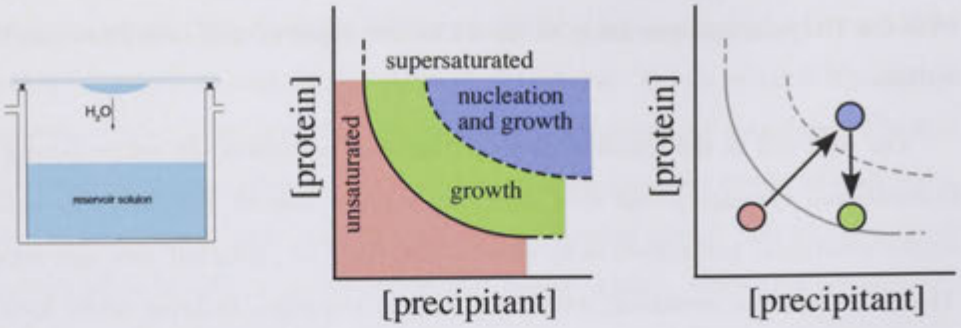


Figure 2.6
On the left is a schematic of the popular 'hanging drop/vapour diffusion' method of growing protein crystals, showing the changes in precipitant concentration as a result of vapour diffusion. The middle and right hand side show a typical phase diagram, and describe the concurrent changes in protein and precipitant concentration during crystal formation respectively (adapted from *'Crystallography Made Crystal Clear'* [39], with permission).

additives in a small drop, and then allowed to equilibrate with a reservoir of a larger volume within a closed system. This results in the lower concentration of additives in the drop slowly equilibrating with the solution of higher concentration in the reservoir via the vapour diffusion of water. This slow increase in protein concentration leads to a decrease in solubility, at which point there are typically two possibilities: the protein can come out of the solution as an amorphous precipitate, or as an ordered protein crystal. For the latter possibility to eventuate, additive concentration needs to equilibrate within a range suitable for crystal nucleation. As the protein crystals form, the concentration of soluble protein should ideally fall into a range optimal for crystal growth, resulting in crystals suitable for X-ray diffraction. This process is often represented with 'phase diagrams', as shown in Figure 2.6.

2.5.2 Data collection and Fourier methods

When X-rays interact with electrons in a molecule, the fluctuating electromagnetic field of the incident radiation forces them to oscillate at the same frequency. The radiation that is generated from this interaction is of the same wavelength, but precisely 180° out of phase with the incident wave. This is known as coherent scattering and occurs in all directions. When monochromatic X-rays interact with molecules within a crystal, coherent scattering also occurs in all directions, but the symmetrically ordered lattice means that the many secondary waves are superimposed on each other. When waves superimpose such that their peaks and troughs align, *i.e.*, in phase, the amplitude of the wave increases in a

Figure 2.7
Shown on the left, coherent scattering of X-rays by adjacent atoms and its relationship to Bragg's law (Equation 2.1) on the right. This states that two waves scattered by a set of planes must have a difference in path length equal to an integral number of their wavelength in order to satisfy the requirements for constructive interference, where λ = wavelength, d = interplanar spacing of atoms and θ = the angle of reflection. (Adapted from http://en.wikipedia.org/wiki/Bragg's_law).

cumulative fashion (constructive interference). When these waves superimpose such that their peaks and troughs are out of phase relative to each other, they effectively cancel each other out (destructive interference). This phenomenon is known as diffraction and results in a pattern of discrete spots when used to expose photographic film or a modern day CCD (charged coupled device) detector. These spots correspond to waves whose intensity have been enormously amplified by the process of constructive interference, and must exit from the crystal at an angle that satisfies Bragg's law:

$$n\lambda = 2d_{hkl}sin\theta \tag{2.1}$$

Diffracted waves can be thought of as 'reflections' off of planes of electron density within the crystal. Each can be assigned a set of Miller indices describing what plane each was diffracted from. If a unit cell has dimensions a, b and c, and a diffracted wave has Miller indices of h,k,l, it originated from the set of planes that intercept the unit cell h times at intervals of a/h, k times at b/k and l times at c/l. Bragg's law states that for constructive interference to occur, the diffraction angle (θ), and interplanar spacing (d) must be such that the differential path length of two superimposing waves ($2d\sin\theta$), is equal to an integral number of wavelengths ($n\lambda$). The basis of this mathematical relationship is shown in Figure 2.7.

Another way of appreciating Bragg's law is from the perspective of reciprocal space. The diffraction of a lattice is also a lattice, but has dimensions that are inversely proportional to that from which it originated. This is the physical relationship between the unit cell of a crystal and it's diffraction pattern. Distances within the unit cell measured in

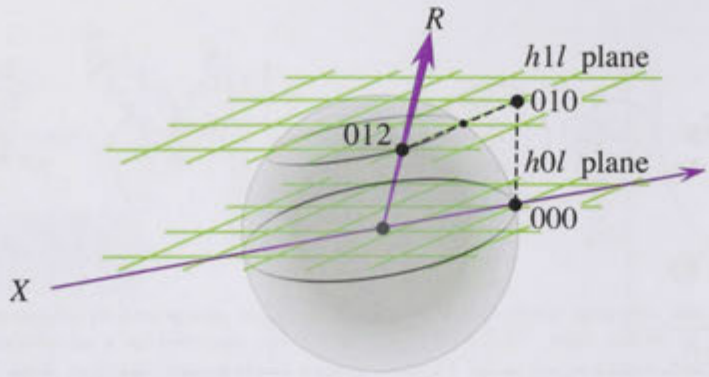


Figure 2.8
Ewald's sphere shown intersecting the reciprocal-lattice planes *h0l* and *h1l*. Reciprocal lattice point 012 is shown contacting the sphere, concurrently producing a ray *R*, at an angle of 2θ relative to the impingement beam of X-rays *X*, that will be recorded as reflection 012 on the detector. Its direction is defined by drawing a line between the origin (000) and reciprocal lattice point 012, that has a length of $1/d_{012}$ and a direction normal to the 012 plane of electron density. (Adapted from '*Crystallography Made Crystal Clear*' [39], with permission).

Å correlate to distances of reciprocal angstroms ($Å^{-1}$). Thus, this lattice is referred to as the reciprocal lattice, and it is composed of measurable reflections in reciprocal space. The position of these reflections in reciprocal space are defined by lines drawn from the origin of the crystal lattice, normal to their corresponding planes of electron density, and of a length $1/d_{hkl}$. It can be shown that the geometric conditions required for a reflection to be measured in reciprocal space are exactly the same as those required to satisfy Bragg's law in real space. A sphere of radius $1/\lambda$, representing the wavelength of X-rays in reciprocal space, generates a geographical construction by which it is satisfied. As the crystal lattice is rotated about its origin, so too is the reciprocal lattice. As points of the reciprocal lattice contact the sphere, Bragg's law is satisfied, and a reflection occurs. This construction is known as the *sphere of reflection* or *Ewald's sphere*, and is demonstrated graphically in Figure 2.8. This implies that the position of reciprocal lattice points are entirely dependent on the size and orientation of the unit cell, and not on its contents. As will be shown below however, it is the *intensities* and *phases* of these reflections that give us the structural information we seek.

The *Fourier transform* is a mathematical function that describes the relationship between a crystal and its reciprocal lattice precisely. It converts the complicated 3-dimensional electron density function into its many component wave functions. Of most importance, it is a reversible operation, such that a reciprocal lattice can be used to calculate the crystal lattice parameters, and vice versa. Fourier demonstrated that for any function $f(x)$, there exists another function $F(h)$, who share the relationship:

$$F(h) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i(hx)} dx, \tag{2.2}$$

where $F(h)$ is the Fourier transform of $f(x)$, i is the imaginary number $\sqrt{-1}$, and the units of the variable h are the reciprocal of the units used for x. For convenience, the phases are implicitly specified in Equation 2.6 by the following relationship of complex number theory, known as Euler's formula:

$$\cos\theta + i\sin\theta = e^{i\theta} \tag{2.3}$$

Conversely, the same mathematical transformation that created $F(h)$ from $f(x)$, can also be performed in the opposite direction:

$$f(x) = \int_{-\infty}^{+\infty} F(h) e^{-2\pi i(hx)} dh \tag{2.4}$$

In the context of X-ray diffraction from crystals however, Fourier's relationship becomes slightly more complicated. Firstly, the diffracted rays responsible for spots in a diffraction pattern are not simple X-ray waves, but *structure factors*, each a Fourier summation of the coherent scattering from every atom in the crystal. While the orientation and dimensions of the unit cell dictate the positions of the reflections, the detailed information about its contents are encoded in the *phase* and *amplitude* of each. Secondly, the above equations pertain to 1-dimensional functions, while the electron density within a crystal is 3-dimensional. Similarly, the terms involved in it's Fourier transform are also 3-dimensional:

$$F_{hkl} = \int_V \rho(x,y,z) e^{2\pi i(hx+ky+lz)} dV \tag{2.5}$$

Here, each structure factor, F_{hkl}, is a summation of contributions from each volume element, V, within the unit cell. The electron density, $\rho(x,y,z)$, is in turn the transform of the structure factors:

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} F_{hkl} e^{-2\pi i(hx+ky+lz)} \tag{2.6}$$

Assuming that the variables within Equation 2.6 are accurately known, calculating electron density maps is therefore a simple process of vector summation. Unfortunately, this is not possible because both the amplitude *and* phase of the X-rays are required, and

only the former is currently able to be measured experimentally. A detector capable of measuring phase would require a sensitivity high enough to differentiate the undulating wave intensity of a coherent X-ray source on a time scale considerably less than the wave's time period (time taken for the wave to travel one wavelength). For X-rays possessing a wavelength of 1 Å, this can be calculated:

$$since\ T = {}^{1}/_{f}\ and\ c = f\lambda,\ \therefore T = {}^{\lambda}/_{c} = \frac{10^{-10}\ m}{3.0 \times 10^{8}\ m \cdot sec^{-1}} = 3.3 \times 10^{-19} seconds \quad (2.7)$$

where T is the time period of the wave, f is the frequency of the wave and c is the speed of light. Since the corresponding time scale of data collection is minutes to hours when using conventional X-ray sources, or seconds when using synchrotron radiation, this is clearly not possible. All phase information is lost when X-rays are accumulated for such long periods of time. This unknown variable in the measurement of experimental diffraction data is commonly known as the *phase problem*.

2.5.3 The phase problem and molecular replacement

When solving crystal structures that are isomorphous with a known structure, the calculated phases from the known structure can be used directly as initial estimates. For instance, when performing experiments with crystals of the previously published trigonal space group of hGSTO1 [4], phases could be obtained from the known contents and orientation of the asymmetric unit immediately. The *observed* amplitudes (proportional to the square root of measured intensities) of the experiment were simply combined with *calculated* phases from a Fourier transform of the known structure. Even though these phases were not accurate, they were close enough to calculate interpretable electron density and hence, solve the structure. The information encoded within the measured amplitudes were then sufficient to reveal any minor differences between the two, such as ligand binding. This concept is eloquently demonstrated by Kevin Cowtan's 'fourier cats' (Figure 2.9 on the facing page) [39].

In Figure 2.9 on the next page, the cat and the manx cat are in identical positions and orientations. If they were not, the cat would first need to be superimposed onto that of manx cat before its calculated phases could be used as estimates. This superposition is known as molecular replacement and was used to solve the structures of the hGSTO1△155 isoform (Appendix A on page 181), as well as the two novel crystal forms of wild type

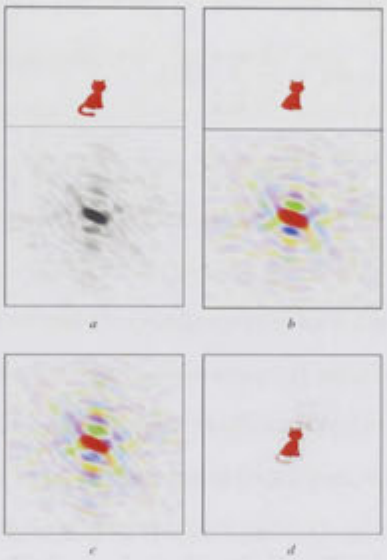


Figure 2.9
The 'Fourier cats' of Kevin Cowtan (http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html). A tailed cat is shown with it's Fourier transform, but is devoid of phase information (grayscale image) (a). Similarly, a 'manx' (tailless) cat is shown with it's Fourier transform, containing intact values of phase (different values within the colour spectrum correspond to different values of phase) (b). This is an analogy of the situation often faced in crystallography during isomorphous replacement and molecular replacement. The question is, are the amplitudes powerful enough to reveal differences in the new structure when combined with phases from the old? By performing a reverse transform combining the manx phases with the cat amplitudes (c), we can see that they are. The tail, albeit at partial occupancy, is clearly visible (d). (Adapted from '*Crystallography Made Crystal Clear*' [39], with permission).

hGSTO1 identified during this research (Chapter 6 on page 131). This involved searching for the orientation and position of the asymmetric unit within the new crystals by using the previously solved structure (PDB identifier: 1EEM), expected to be similar. Although the problem could theoretically be approached by calculating the dimensions and orientation of the new unit cell from the raw diffraction data, then calculating phase estimates at every conceivable position, the required computational time is decreased dramatically by separating this complicated search into two separate searches of orientation *and* position. This can be achieved by using the Patterson function. Mathematically, it has the general form:

$$P(u,v,w) = \frac{1}{V}\sum_{hkl} |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)} \tag{2.8}$$

Since the Patterson function contains no phases, it can be computed from any set of raw crystallographic data, and because the previously solved structure was expected to be very similar to that in the new crystals, it follows that they would have very similar Patterson maps. In contrast to a contour map of electron density that displays high peaks of density that correspond to atoms (x, y, z), the Patterson function creates a contour map that displays peaks corresponding to *vectors between atoms*, (u, v, w). There are many more vectors between atoms than there are atoms, so Patterson maps are more complicated. Conceptually, an arrow (vector) is drawn between every atom and every other atom in a protein molecule. The tail of each is then moved to the origin to create the Patterson map. Since the distances between atoms in a protein are independent of the proteins *position*, it is only the *orientation* that determines the appearance of the Patterson function. As the structure rotates, so does the Patterson map. This can be used to find the *best orientation* of a protein within a new unit cell. Using the results of this *rotation function*, a separate search can then be performed to find the *best position*. This is achieved using a *translation function*, in which the optimally orientated model is translated about the unit cell in real space. Intensities are iteratively calculated and compared with those observed in the diffracting pattern. At a certain position, these differences will be at a minimum. This can be quantified using the crystallographic R-factor (Equation 2.13 on page 41), correlation coefficient or maximum likelihood function. The resulting reduction in computational time and effort required to find a solution is enormous.

2.5.4 An introduction to anomalous scattering

The electrons orbiting the nuclei of atoms do so within specific orbitals. Each of these orbitals possess a specific energy. In order for electrons to pass from one to another, energy must be emitted or absorbed. As energy of the incident radiation increases, so does the absorption of the element, until it sharply drops away. This is known as the elements *absorption edge*. Just above this value, there is sufficient energy to eject inner shell electrons from their orbitals. This results in X-rays of a specific wavelength being emitted as outer shell electrons fall down to take their place. This wavelength directly corresponds to the gap in energy between the two orbitals.

At synchrotrons, the wavelength of X-rays can be finely tuned such that it corresponds to the absorption edge of heavy elements such as arsenic (Figure 2.10 on the next page), but not lighter elements such as C, N and O, due to their absorption edges not occurring within the range of X-ray radiation useful for crystallography. This results in radiation being absorbed and re-emitted with an altered phase. This is known as *anomalous scattering* or *anomalous dispersion*. As crystal lattices are inherently symmetrical, so are their diffraction patterns. The principle effect of anomalous scattering is the break down of this symmetry relationship, known as *Friedel's law*. This law states that a particular reflection, with Miller indices hkl, will have an intensity equal to the reflection with the same indices, but with opposite sign, $\overline{hkl}$. One can think of each of these reflections coming from the same planes of electron density at the same angle, but from opposite sides. Friedel's law is broken because the phase contribution of absorption and re-emittance is symmetrical in its real component, but not in its imaginary component (Figure 2.10 on the following page). This is demonstrated with vector diagrams in Figure 2.11 on page 35.

Subtraction of native and anomalous intensities dramatically simplifies the Patterson function (Equation 2.8 on the facing page) to just a few atoms. This very simple 'structure' is easy to solve, subsequently allowing the position, and thus the phase, of the heavy atoms to be calculated (Equation 2.6 on page 29) [40]. Therefore, an electron density map can be calculated (Section 2.5.5.2 on page 37) of just these atoms, but in the absence of a molecular replacement model, can also be used to generate initial phase estimates for the whole protein as well. Since *all* atoms in the molecule contribute to the generation of *all* structure factors, the quantifiable contribution of anomalous dispersion allows an initial estimate of phase to be calculated for each. In practice, this requires several datasets collected at wavelengths with varying anomalous dispersion contributions. Often 3 datasets

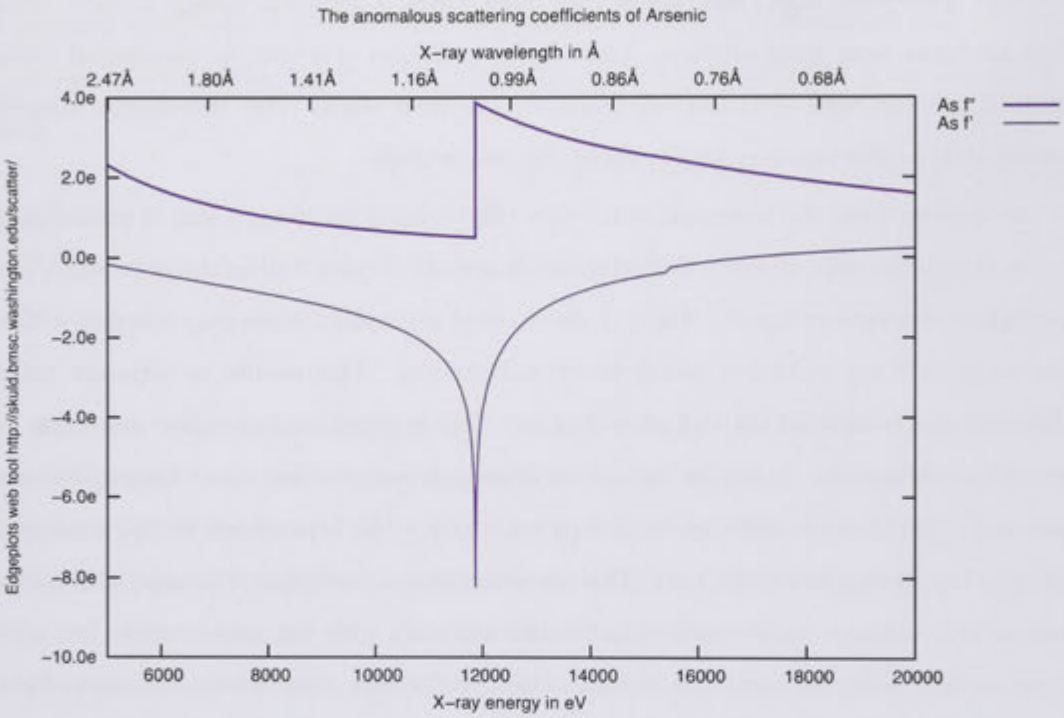
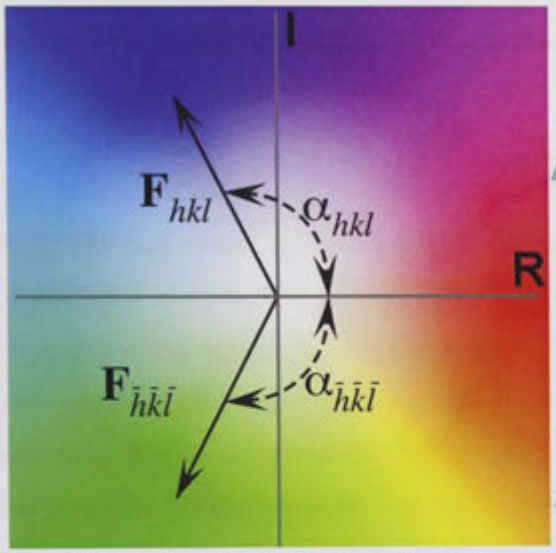
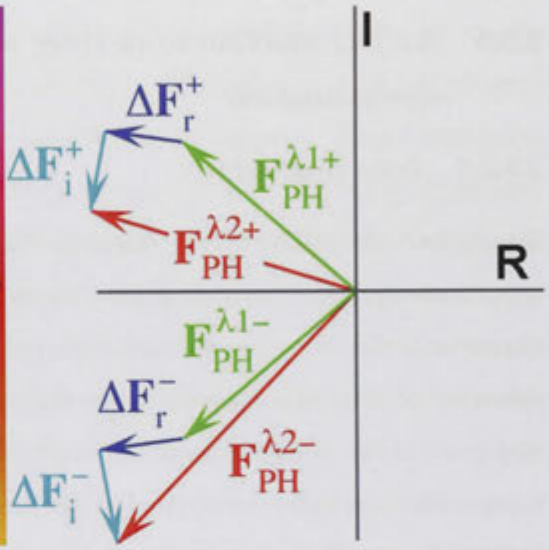


Figure 2.10

Due to its application in this research, the theoretically determined real (f') and imaginary (f'') anomalous scattering coefficients of Arsenic are shown to demonstrate the physical basis of anomalous dispersion. Experimentally, the imaginary component of this phenomenon is directly proportional to the observed X-ray absorption and fluorescence. Once the f'' spectrum has been determined, it can be used to calculate that of f'. This wavelength dependent contribution to structure factor phase and amplitude can be exploited by the variable wavelength of synchrotron radiation to locate the position of the atoms responsible. Subsequently, this information can be used to calculate phase estimates for the entire structure. The methodology is demonstrated most eloquently by representing structure factors as vectors on the complex number plane (Figure 2.11 on the next page). This figure was generated using the 'edge plots' tool at the University of Washington web server (http://skuld.bmsc.washington.edu/scatter/).



(a) The structure factors of a Friedel pair represented on a complex number plane. The colour spectrum represents the corresponding phase value.



(b) At wavelength λ_2, F_{hkl} is no longer the mirror image of $F_{\overline{hkl}}$ due to anomalous scattering. While the phase shift of anomalous scattering is mirrored in its real component (ΔF_r), this is not the case for its imaginary component (ΔF_i). The resulting asymmetry in phase leads to an asymmetry in amplitude. Measuring the same reflection using λ_1, in the absence of anomalous dispersion, allows its contribution to be clearly visualised by simple subtraction. This quantifiable contribution of the anomalous diffraction to each structure factor, coupled with the knowledge of the positions of the atoms responsible, allows a phase estimate to be calculated for each.

Figure 2.11

A depiction of Friedel's law on the complex number plane in the absence and presence of anomalous dispersion. The x and y axis represent the real and imaginary number planes. Each structure factor is represented as a vector, with its amplitude and phase indicated by the vector length and angle formed with the real axis, respectively. The real and imaginary contributions of anomalous dispersion can be quantified via head to tail summation. (Adapted from '*Crystallography Made Crystal Clear*' [39], with permission).

are collected in order to adequately overdetermine the phase problem: one at the peak of f" absorption, another at the peak of f' absorption, and a native dataset at a wavelength far from the absorption edge. This technique is known as multiple-wavelength anomalous dispersion (MAD) [41], and is the most commonly used technique in the structural determination of novel protein folds.

2.5.5 An introduction to *in silico* methods of macromolecular structure determination

2.5.5.1 Data processing

Throughout the course of this research, the HKL2000 package [42] was invariably used for processing of the raw diffraction data, each a 2-dimensional projection of a curved 3-dimensional slice of reciprocal space. The process of reducing single crystal diffraction-data into a list of indexed reflections can be broken down into several steps. These can be summarised: (1) the visualisation of the original, unmodified detector data, (2) assigning the many reflections with a corresponding *hkl* index, (3) refinement of the crystal and detector parameters to reflect the experimental conditions, (4) integration of the diffraction intensities, (5) applying a scale factor to equivalent reflections to account for variables such as crystal thickness, (6) a global precise refinement of crystal parameters using all recorded reflections, and (7) the statistical analysis and merging of reflections related by crystal symmetry. The HKL2000 package includes the Denzo and XDisplayF programs to index and integrate the reflections (steps 1-4), while the scaling of equivalent reflections into a single list of indexed intensities (steps 5-7) is performed by the program Scalepack. The first step in indexing is to perform a 'peak search' within XDisplayF. This identifies reflection peaks in a single diffraction image. Denzo then uses a novel algorithm that combines *a priori* knowledge about the experimental geometry to map them into reciprocal space. The program performs a search of all possible indices of all reflections, and exploits the ability of each hkl index to be determined independently of the other two. This ultimately allows the possible unit cell size and orientation to be determined. The crystallographer is then given a list of the Bravais lattice symmetries (Section 2.5.1 on page 24) coupled to a distortion index. As a general matter, the lattice of highest symmetry which displays a modest distortion index will be a clear outlier given good data. A trial and error process of comparing the merging statistics of symmetry related reflections when scaled can then ascertain the correct choice of space group. Calculation of initial phase estimates of the in-

tegrated and scaled reflections can then occur, for e.g., via MAD (Section 2.5.4 on page 33) or molecular replacement (Section 2.5.3 on page 30) experiments discussed previously.

2.5.5.2 Calculation of electron density maps

Due to the calculated phases being only rough estimates, the initial electron density maps can be somewhat uninformative. Crystallographers improve these phases by iterative methods, sometimes referred to as 'bootstrapping', whereby identifiable features in the electron density are successively included in the next round of map calculation. Within real space, the major method of phase improvement involves altering the model to fit the experimentally determined electron density. The F_{obs} electron density map, calculated by combining the observed amplitudes with the calculated phase estimates from the search model, may initially only show a rough outline of the molecular envelope.

The function is first modified to reflect known properties of protein and water within crystals. Knowing protein crystals generally contain about 50% water, the ratio of high to low density can then be adjusted to agree with that of protein to solvent. Commonly used techniques include 'solvent flattening', 'skeletonisation' and 'histogram matching'. Solvent flattening involves an algorithm that first defines the molecular envelope occupied by protein, and then averages out aberrations in the remaining unit cell contents to reflect the homogenous diffractive qualities of solvent. Skeletonisation then attempts to trace a continuos 'tube' of electron density within the envelope corresponding to the polypeptides main chain. These atoms are generally the most ordered and therefore expected to yield the most interpretable electron density. Regions of positive density not within a reasonable bonding distance of this main chain model are then also down weighted to reflect the expected density of solvent. Histogram matching then matches the electron density function of successive atoms to resemble those of high resolution crystal structures in the PDB. This manipulated function describing the electron density is then input into Equation 2.5 on page 29 to compute structure factors. These updated phases are then combined with the observed amplitudes for the next round of map calculation using Equation 2.6 on page 29. The old phases are not thrown out straight away, but are weighted in terms of their reliability, so as not to overly bias the data. If the new phase estimates are indeed better, the molecular envelope will be clearer and the cycle is repeated. If there is non-crystallographic symmetry present (NCS) , the map may also be improved by averaging the density over the different molecules within the asymmetric unit, effectively increasing

the signal to noise ratio and facilitating a clearer image of the molecule. Finding NCS within an asymmetric unit is another application of the Patterson map, referred to in this context as a *self-rotation function.*

At some point in the phase improvement process, the map becomes clear enough to locate specific features of the model such as the shapes of amino acid side chains, which can then be used to calculate atomic structure factors instead:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)} \tag{2.9}$$

Conversion from a density based model to one based on the coordinates of an atomic model greatly increases the number of refinable parameters however. If introduced too early, there is a risk of biasing the refinement process. This is especially likely when obtaining phases from molecular replacement, in which calculated phases from the phasing model may disproportionately influence the electron density. This effect is known as *phase bias.* Although the observed amplitudes of structure factors are capable of restoring missing information, it can be easily obscured by the inaccuracies within their calculated phase. A randomly assigned phase has a more profound effect on the electron density function than a randomly assigned amplitude does. If not detected, the crystallographer may mistake vestigial features from the phasing model as real, further reinforcing this error in subsequent calculations. It is therefore advantageous to use Fourier syntheses that minimize this bias as much as possible. Amplitudes in the Fourier sum of $F_{obs} - F_{calc}$ and $2F_{obs} - F_{calc}$ are the most common. Using the $F_{obs} - F_{calc}$ coefficients as an example, electron density could be calculated:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} (|F_O| - |F_C|) e^{-2\pi i(hx + ky + lz - \alpha'_{calc})} \tag{2.10}$$

where F_O are the observed structure factor amplitudes, and F_C and α'_{calc} are the amplitudes and phases calculated from the model. Maps created using these types of Fourier syntheses are termed $F_O - F_C$ and $2F_O - F_C$ maps respectively, and give the observed amplitudes the greatest chance possible of guiding the crystallographer to the correct structure. While the latter creates a 3-dimensional contour map into which the molecular model can be built, the former allows minor differences between the amplitudes of the calculated structure factors and those observed in the diffraction pattern to be visualised. The presence of a structural feature in the model that was not corroborated by the observed structure factors would

be reflected as a negative peak in the $F_O - F_C$ map, whereas missing structural features in the map not accounted for by the model would create a positive peak. Often structural aberrations such as an incorrect amino acid side chain rotamer assignment, or the need for a main chain carbonyl flip, can be identified by the placement of a negative peak associated with the model, immediately adjacent to a positive peak over a neighbouring region of unfilled space. In this way, the working model can then be adjusted slightly until local aberrations in the $F_O - F_C$ map subside, at which point the observed intensities match those calculated from the model almost exactly. The $2F_O - F_C$ map on the other hand is usually all positive, and gives the crystallographer a much more intuitive and interpretable guide to correcting the structure. Note that in Figure 2.9 on page 31, the missing feature of the tail is only restored at half occupancy. This only occurs when the phase estimates are almost correct. The factor of a half arises from the observed amplitudes only correcting the structure factor in the direction parallel to the calculated phase. Due to the error in it's phase angle however, it also requires correction in the perpendicular direction, and this correction is statistically likely to be equal in magnitude. This suggests that one can double the observed amplitudes and subtract from this those that are calculated. The resulting map should show features in the current model at full weight, but should also show the multiplied information from missing structural features at full weight as well. Some crystallographers prefer a $3F_O - 2F_C$ map for this purpose, a compromise between the two. Using the standard deviation and maximum likelihood methods, weighting these map coefficients is also commonplace, to further guard against the introduction of bias. The equivalent to $F_O - F_C$ maps in this case is $mF_O - DF_C$ maps, where m is a figure of merit for each phase calculated from the model, and D is an overall estimate of atomic coordinate error. When the model has been fitted to the map to the best of their ability, specialized software can be used to further improve the model within reciprocal space. In some applications of this Fourier synthesis, D is replaced by σ_A. In this instance, D has been corrected for completeness of the current model.

2.5.5.3 Model building and refinement

Once the phases have become accurate enough to allow the building of a near atomic model, interactive computer graphics can be used to manually place aberrant parts of the model into its respective electron density. These programs also contain numerous other features to assist in model building, such as 'real-space' refinement of amino acids into

electron density by the method of least squares, and a 'regularize' function, that imposes ideal bond lengths and angles to freshly rebuilt regions of the polypeptide. Throughout the course of this research, the program Coot [43] was employed for this purpose.

When the crystallographer has fitted the model to the electron density to the best of their ability, the structure can be subsequently refined in reciprocal space. Due to the number of parameters to be fitted usually outnumbering the number of observations (reflections) recorded in macromolecular crystal diffraction, the refinement software also enforces realistic bond length and geometry constraints based on those of ultra-high resolution crystal structures. This effectively increases the 'observations', allowing the refinement process to be sufficiently overdetermined. Mathematically, the process of reciprocal space refinement minimises a function Φ, whose form is:

$$\begin{aligned} \Phi &= \sum_{hkl} w_{hkl}(|F_O| - |F_C|)^2 hkl \\ &+ \sum_{i}^{bonds} w_i(d_i^{ideal} - d_i^{model})^2 \\ &+ \sum_{j}^{angles} w_j(\phi_j^{ideal} - \phi_j^{model})^2, \end{aligned} \tag{2.11}$$

where d_i is the length of bond i, and ϕ_j is the bond angle at position j. w_{hkl} is a weighting term that reflects the accuracy to which these variables are known. Statistically, this would typically be defined as $1/\sigma^2_{hkl}$, where σ_{hkl} is the *standard deviation* from multiple measurements of $|F_O|$. Since a crystallographic dataset often does not contain sufficient measurements to calculate a standard deviation, other weighting schemes have been devised. This function simultaneously minimizes the differences between (1) the observed and calculated reflection amplitudes, (2) model bond lengths and ideal bond lengths, and (3) model bond angles and ideal bond angles. Perturbations of parameters that make the model less realistic are therefore penalized. The parameters refined initially are tightly constrained by ideal geometries. As phase estimates become more accurate however, the w_i and w_j terms become smaller while the w_{hkl} term tends towards unity, gradually relaxing constraints on the structure as more experimental parameters are included in Φ. These include variables such as the *temperature factor* B_j, or B-factor of each atom j (also known as the atomic displacement parameter), that quantifies the extent to which an atom oscillates about its specified position in the model due to thermal motion. Another is the *occupancy* n_j of each atom j, that reflects the percentage of the crystal lattice occupied

by a group of atoms. Flexible side chains such as that of glutamate can often adopt alternate conformers within the crystal lattice. This phenomenon is most prevalent at the solvent interface of the polypeptide or in the vicinity of ligand binding. The term F_C in Equation 2.11 on the preceding page can thus be expanded:

$$F_C = G \cdot \sum_j n_j f_j e^{2\pi i(hx_j+ky_j+lz_j)} \cdot e^{-B_j[sin\theta/\lambda]^2}, \tag{2.12}$$

where G is an overall scale factor to convert all F_Cs to a convenient value, n_j is the occupancy of atom j and B_j is its temperature factor, whose effect is proportional to the angle of reflection $[\sin\theta/\lambda]$.

As refinement proceeds, the agreement between the model and the raw diffraction data is commonly quantified with the *residual index*, or R-factor:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \tag{2.13}$$

The R-factor essentially compares calculated amplitudes with those observed. The lower the R-factor, the better the agreement. To avoid 'overfitting' the data (refining more experimental parameters than there is experimental observations), the free R-factor is also employed, generally calculated from 5-10% of the measured reflections that are *not* refined against. This independently tests the ability of the model to predict reflection amplitudes that were not used to guide its refinement during previous Fourier transforms, and is generally regarded as a less biased quantifier of model accuracy.

The structures presented in this dissertation where refined using the Refmac [44] and Phenix [45] programs, that both employ the *maximum likelihood* target function during refinement. The latter even has a 'simulated annealing' option, which simulates heating the structure up and then cooling it down, allowing it to escape from local energy minima introduced by factors such as phase bias.

This process of reciprocal and real space refinement of the structure was repeated until the reduction in R and R-free values converged and in most places, the model fit the calculated electron density precisely. At this stage, the structure was validated using the program MolProbity [46], or the comprehensive validation tools inbuilt into Phenix [45].

The complex of hGSTO1 with ascorbate:

An active site binding mode suggestive of product inhibition

Due to the considerable instability of the substrate dehydroascorbate, the product ascorbate was alternatively employed in crystallisation. It was aimed to discover the binding mode of the product at the conclusion of the catalytic cycle. One such experiment has allowed the binding of this antioxidant to be characterised in the active site of hGSTO1, but in a mode indicative of a product inhibition mechanism. The implications for its proposed mechanism of enzymatic reduction are discussed.

3.1 Introduction

Although the reduction of oxidized vitamin C by hGSTO1-1 was initially reported over a decade ago [4], the expression and subsequent characterisation of hGSTO2-2 was initially problematic. Recent developments have since allowed its biochemical activity to be defined, and it's activity to this substrate was found to be approximately 100-fold greater than that of its previously characterised counterpart. In addition, a survey of published mammalian dehydroascorbate reductase (DHAR) activity shows its activity to be far in excess of enzymes found in other mammals, suggesting this may be the main role of hGSTO2-2 within the body [25]. hGSTO2 shares a high level of sequence similarity with hGSTO1, including the active site cysteine residue at position 32 [14]. While the activity of hGSTO1-1 towards this substrate is lower than that of it's isozyme, the chemical mechanism and mode of binding is likely to be highly conserved between the two. Based on a previous characterisation of this reaction in the context of a mammalian glutaredoxin [47], the mechanism shown in

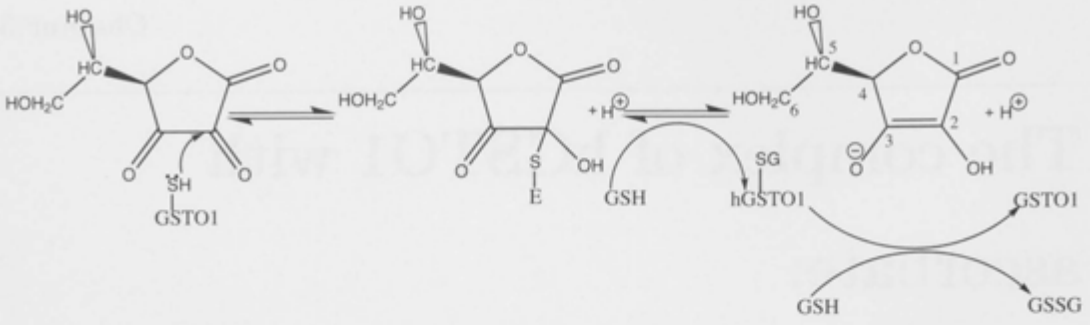


Figure 3.1
A proposed mechanism of dehydroascorbate reduction by hGSTO1-1, based on the characterisation of this rection when catalysed by pig liver thioltransferase (glutaredoxin) [47].

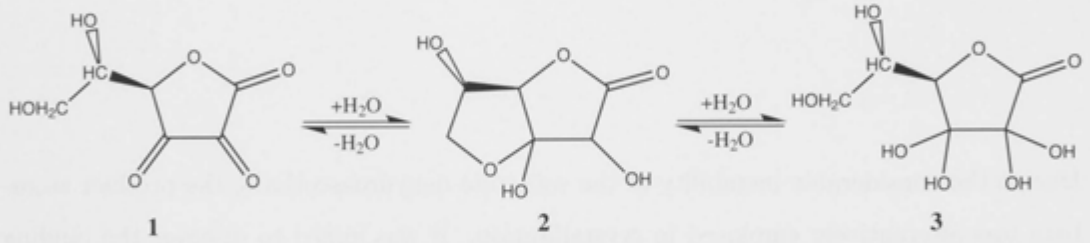


Figure 3.2
The hydration equilibria of DHA, in which the bicyclic hemiacetal hydrate **2**, has been shown to predominate [48].

Figure 3.1 could be proposed.

While the 1,2,3-tricarbonyl structure of dehydroascorbate (DHA) (left hand side of Figure 3.1) is the chemical form reported in many textbooks, the true nature of this species in solution is somewhat more complicated. Quantum mechanical calculations suggest that the tricarbonyl would only be stable for fractions of a second, and combined with evidence from crystal structures of various derivatives, and the H^1 and C^{13} NMR spectroscopy of this compound in solution, the most representative structure of this species is presently thought to be the hemiacetal hydrate **2** [48], shown in Figure 3.2.

The conversion of ascorbate to this bicyclic intermediate is an inherently multistep process that requires the loss of two electrons and two protons. The principal intermediate of this reaction is the radical species semi-dehydroascorbate (SDAH), $C_6O_6H_6^{\cdot -}$. Although the precise structure of this anion-radical is unknown, it primarily participates in a disproportionation reaction to give ascorbate and DHA. Kinetic analysis suggests that this mechanism features a dimer of unknown structure that participates in a reaction with hydrogen donors to yield the products [48].

For clarity, DHA and its derivatives will continue to be represented in the tricarbonyl

form in subsequent discussions of mechanism. However, the reader is asked to keep in mind the complex interaction of cyclisation, radicalisation and dimerisation that would concurrently occur.

Convergent genetic mapping has implicated the Omega class GSTs in early onset of two neurodegenerative conditions, Alzheimer's and Parkinson's disease [49, 22], and hGSTO1-1 has been shown to be the target of anti-inflammatory drugs affecting cytokines also implicated in their pathogenesis [24] (Chapter 6 on page 131). This is provocative given that ascorbate is suspected to play a key role in protection from the oxidative stress implicated in the pathogenesis of these conditions. This may be of particular importance in the brain where its concentrations is most tightly regulated [50].

Soaking and co-crystallisation experiments with native hGSTO1 has resulted in non-diffraction and a failure to crystallise respectively. Experiments involving crystals of the C32A and C32S active site mutants were therefore employed, hoping that the resulting enzymatic inactivity would stabilise the binding mode of the product within the hGSTO1 active site, allowing it to be visualised. Here, the cysteine 32 to serine active site mutant is presented in a high resolution crystal structure. Although an analogous binding mode is expected within the C32A mutant, no crystals of this isoform having been soaked or co-crystallised with the compound diffracted to a sufficient resolution to enable this to be confirmed.

3.2 Experimental Protocols

The C32S active site mutant first needed to be made and cloned into the pHUE vector. After transformation, expression and purification, hGSTO1 was co-crystallised with L-ascorbate. Further details of each experimental step are outlined below.

3.2.1 Site directed mutagenesis

Mutagenic primers of 23 base pairs in length were designed with a suitable melting temperature and obtained from GeneWorks Pty Ltd. KOD Hotstart polymerase (Novagen) and Pfu Turbo polymerase (Promega) were then used to perform multiple C32S mutagenic polymerase chain reactions (PCR) (Figure 2.1 on page 17) upon pGEM-T-*hGSTO1* template. The methylation specific endonuclease, *Dpn1* (NEB), was subsequently employed to digest template DNA before transforming electroporation competent *E. coli* cells of strain DH5α with the mutagenic plasmids. These were immediately used to inoculate LB

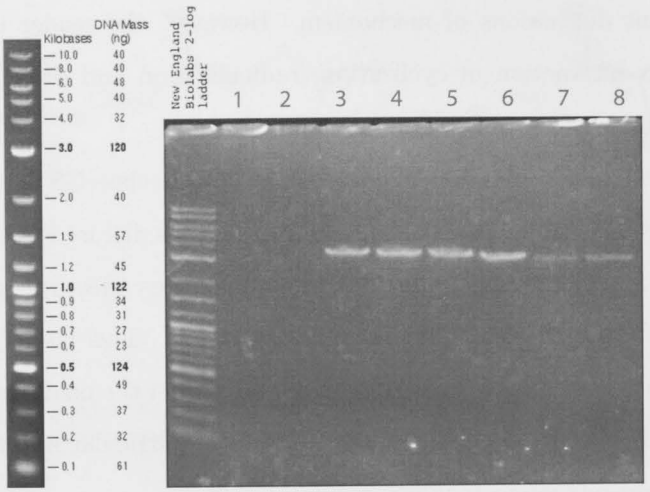


Figure 3.3
An ethidium bromide agarose gel visualisation of site directed mutagenesis products, post *Dpn1* digestion. All even numbered lanes were loaded with PCR reactions containing 4 ng of pGEM-T-*hGSTO1* template, all odd numbered lanes were loaded with reactions containing 2 ng. The PCR products loaded in lanes 1,2, 7 and 8 were performed using Pfu Turbo polymerase (Promega), those in the remaining lanes used KOD Hotstart polymerase (Novagen) instead. 5 μL of each 50 μL reaction was loaded in each lane. 1 μL of NEB 2-log DNA ladder was loaded in the lane on the far left. A reference figure indicating the size of each band in the ladder is shown adjacent.

agar plates (0.2 mg/ml Ampicillin) that were incubated face down overnight at 37°C. The next morning, single colonies were used to inoculate 10 ml cultures of LB (0.2 mg/ml Ampicillin). Subsequent to a 'MiniPrep' (Qiagen) purification of plasmid, the success of the reactions was confirmed via separation on a 1% agarose gel in electrophoresis buffer, containing 0.5 μg/ml of ethidium bromide. The gel was subsequently visualised with a UV transilluminator at 366 nm. (Figure 3.3). Plasmid from 6 of the 8 mutagenesis reactions performed was observed at the expected positions relative to the simultaneously loaded '2-log' DNA ladder (NEB). A complementary *Xma*I/*Sac*I (NEB) digestion of this DNA was then used for gel purification of the insert fragment, and subsequent ligation into the pHUE vector as described in Section 4.2.1 on page 66. The desired TGC to TCC (serine) mutation was then confirmed unequivocally by DNA sequencing (Figure 3.4 on the next page).

3.2.2 Purification of hGSTO1 C32S

The purification procedure used to obtain protein for crystallization was as follows. The pHUE-*hGSTO1* C32S plasmid was transfected into *E. coli* cells of strain BL21 via electroporation. Streaking onto an LB agar plate containing 0.2 mg/ml of Ampicillin and incubating face down overnight at 37°C resulted in multiple single colonies. One of these

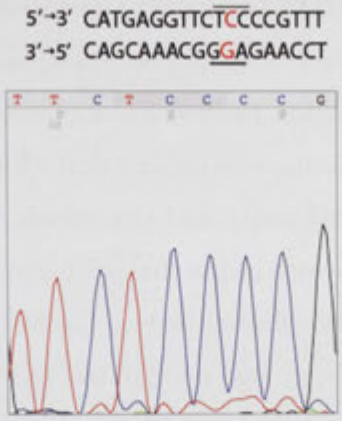


Figure 3.4
The DNA sequencing chromatogram of the cysteine 32 to serine active site mutation unequivocally confirming the TGC (cysteine) to TCC (serine) nucleotide permutation. The forward and reverse primers used to introduce the serine residue are also shown above its sequence. The corresponding trinucleotide is shown with a bar, with the mutagenic nitrogenous bases highlighted in red.

colonies was then used to inoculate a 50 mL starter culture of LB (also containing 0.2 mg/ml of Ampicillin), before subsequent shaking at 37°C overnight. 10 mL aliquots of this starter culture were then used to inoculate five 1 L volumes of LB (0.2 mg/ml Ampicillin) in culture flasks. These were then grown at 37°C with gentle shaking until their optical density reached ≈0.6 at 600 nm. At this point, *hGSTO1* expression was induced via the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 50 μM, and the cultures grown for an additional 3 hours. Cells were subsequently pelleted via centrifugation in a 8K-10 centrifuge (Sigma) at 5000 rpm for 30 minutes using a Nr.12510 rotor. The cellular pellets were then transferred to a large beaker and frozen at -20°C. Except were noted, all subsequent steps were performed at 4°C. The following day, the pellets were thawed in ice, followed by resuspension in Buffer A (300 mM NaCl, 50 mM Sodium Phosphate, 20 mM imidazole 10% Glycerol, pH 7.5) containing 0.3 mM of the protease inhibitor 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) using a magnetic stirrer and bar. Lysis was then mediated via a French pressure cell. Cell debris was pelleted by centrifugation in a RC5C centrifuge (Sorval) at 15 000 rpm for 20 minutes using a SS-34 rotor. Floating nuclear material was removed with a pipette before the supernatant was vacuum filtered through 0.45 μm filter paper. The filtrate was then passed through a pre-packed 5mL His-trap HP (GE Healthcare) Ni-NTA column. An FPLC mediated imidazole gradient from 20-500 mM in Buffer A was then used to elute bound protein. Fractions containing eluted protein were indicated by an absorbance change at 280 nm and were subsequently checked for purity with SDS-PAGE (room temperature). The combined

fractions found to contain hGSTO1-1 in sufficient purity were then dialysed overnight in 5L of Buffer A, using pre-equilibrated 12 kilo Dalton (kDa) molecular weight cut off (MWCO) dialysis tubing. The next day, the purified Ub-hGSTO1-1 fusion product was removed from dialysis. The deubiquitylating enzyme and dithiothreitol (DTT) was then added to final concentrations of 25 μg/ml and 1 mM respectively. The cleavage reaction was allowed to proceed for 2.5 hours with gentle shaking before dialysis in Buffer A to remove the imidazole and DTT (the latter would reduce the Ni^{2+} within the Ni-NTA resin and cause precipitation). A subsequent pass through the Ni-NTA column facilitated selective removal of cleaved Ub, uncleaved Ub-hGSTO1-1 construct and co-purified contaminants. hGSTO1-1, now devoid of an affinity tag, was collected in a high purity and yield in the flow through as judged by SDS-PAGE (room temperature). Protein solutions were then dialysed overnight in a final dialysis buffer of 60 mM NaCl, 20 mM Tris, 5 mM DTT, pH 8 before concentrating to 25 mg/ml using a 10 kDa cutoff Amicon Ultra centrifugal device (Millipore, Cork, Ireland). Protein concentration was finally determined using a NanoDrop 'ND-1000' spectrophotometer (NanoDrop Technologies) via absorbance at 280 nm (room temperature). Molar extinction coefficients were calculated using the Swiss-Prot Proteomics server [51].

3.2.3 Crystallisation, data collection and processing

The C32S active site mutant was crystallised using the hanging drop, vapour diffusion method as previously described (Section 2.5.1 on page 24), with the following exceptions. A precipitant versus pH screen found that optimal crystal growth was slightly different to that previously published [4]: 2.2 M $(NH4)_2SO_4$ and 100 mM Sodium Acetate, pH 4.75. This may have been due in part, to these crystals being grown in the presence of 100 mM L-ascorbate and α-tocopherol succinate (vitamin E) (Table 3.1 on the next page). The latter was included in this experimental design due to reports of it acting as an uncompetitive inhibitor of hGSTO1-1 monomethylarsonate reduction (Chapter 4 on page 61) [52, 53]. As the binding site of methylarsenate species and dehydroascorbate were expected to be the same, it was hypothesised that the tocopherol ester may also be an uncompetitive allosteric inhibitor of hGSTO1-1 mediated dehydroascorbate reduction, and therefore stabilise the enzyme-product complex. The affect of this inhibitor if any, was not observable in the resulting crystal structure.

Cryoprotection was via transfer of the crystal from the crystallization drop, to an equal

Crystallisation	Experimental Parameters of hGSTO1 C32S-ascorbate complex.
Reservoir solution	2.2 M $(NH4)_2SO_4$
	100 mM Sodium Acetate pH 4.75
Crystallisation drop	1 µl hGSTO1-1 C32S
	1 µl 100 mM L-ascorbate
	1 µl Saturated aqueous α-tocopherol succinate.
	1 µl reservoir
Cryoprotectant solution	2 M Li_2SO_4
	100 mM L-ascorbate pH 4.75
	5 mM α-tocopherol succinate in acetone (5% v/v)
	2.5% (v/v) Glycerol
Data collection	
Resolution limits, Å	30.0-1.70 (1.76-1.70†)
Space group	$P3_121$
Unit Cell, Å	a=b=56.92 c=140.51
Total reflections	215 202
Unique reflections	29 768
Completeness, %	99.2 (98.7†)
I/σ	40.9 (4.14†)
Redundancy	7.2 (7.3†)
Refinement	
R_{work}/R_{free}, %	16.9/21.2
rmsd bonds, Å	0.019
rmsd angles, °	1.641
Number of waters	205
Ligands	L-ascorbate, SO_4^{2-}, glycerol, acetate

†, the highest resolution shell (1.76-1.70 Å).

Table 3.1
Crystallisation and structure determination statistics of the hGSTO1 C32S-ascorbate complex.

sized drop of the cryo-solution. Lithium, rather than ammonium sulfate was used, due to its ability to suppress crystal formation at low temperatures. In addition, α-tocopherol succinate was added from a 100 mM acetone solution to avoid precipitation. The resultant 5% concentration of acetone also served to contribute to the cryoprotection of the solution already afforded by the 2.5% concentration of glycerol. Further details of the crystallisation and cryoprotection conditions are outlined in Table 3.1. Data was subsequently collected with X-rays of an energy of 13001.10 electron-volts (eV), using the PX2 protein crystallography beamline at the Australian Synchrotron, Melbourne, Australia.

The diffraction data was indexed, measured and scaled with the HKL2000 package [42]. Phase estimates were calculated from the previously published native structure (PDB identifier: 1EEM) [4]. Subsequent model building and refinement was performed alternatively with Coot [43] and either the Refmac [44] or Phenix [45] programs, respectively.

3.3 Results and Discussion

Seen to be undergoing a π-stacking interaction with phenylalanine 34, the 1.70 Å structure shows ascorbate residing 5 Å from the mutated C32S residue, occupying a portion of the 'G-site' typically inhabited by the γ-glutamate moiety of glutathione [4]. The other side of the carbon heterocycle is accommodated by a hydrophobic interaction with the side chain of L71. The only significant structural perturbation is the alternate conformers of E85. In addition to the rotamer previously interacting with the γ-glutamyl amine of glutathione [4], the binding of ascorbate has evoked its movement by several angstroms with an occupancy of 0.45 (Figure 3.5 on the facing page). The new conformer concurrently engages in polar interactions with the c-6 (for numbering please see Figure 3.1 on page 44) hydroxyl of ascorbate and the side chain of S86. The main chain amide of the latter also forms a hydrogen bond with its c-5 hydroxyl. Other interactions within the active site include a hydrogen bond between its c-3 hydroxyl group and the backbone carbonyl of V72. A secondary interaction between the introduced S32 residue and the c-2 hydroxyl of ascorbate is also mediated by an acetate molecule, observed 2.65 Å from the serine hydroxyl group (Figure 3.6 on page 52).

The observed proximity of the enzymatic product to the active site residue, in this case artificially mutated to a serine, is suggestive that this binding mode is physiologically relevant. It's occlusion of the G-site however, implies this may not be a mode of binding relevant to catalysis, but may instead be suggestive of product inhibition. In the mechanism proposed (Figure 3.1 on page 44), generation of ascorbate is coupled to the formation of a mixed disulfide between the active site cysteine and glutathione. This would be difficult to reconcile with the mode of ascorbate binding observed, since it is the same position observed to accommodate the glutathionyl functionality of this disulfide in the hGSTO1 native crystal structure [4]. Likewise, the observed binding mode of the product is unlikely to be transposable to a preliminary docking position of the substrate, since the removal of conjugation between the 1-carbonyl and the 2,3-alkene within the hemiacetal heterocycle (Figure 3.2 on page 44) destroys its planarity, both of which would make the equivalent π-stacking interaction with the side chain of F34 unfavourable (Figure 3.6 on page 52). The crystal structure might therefore feature within a negative feedback loop, inhibiting the generation of ascorbate when its intracellular concentration becomes too high by blocking entrance to the G-site. Such a mode of binding would have been encouraged in this experiment by the artificially high concentration of ascorbate present (100 mM).

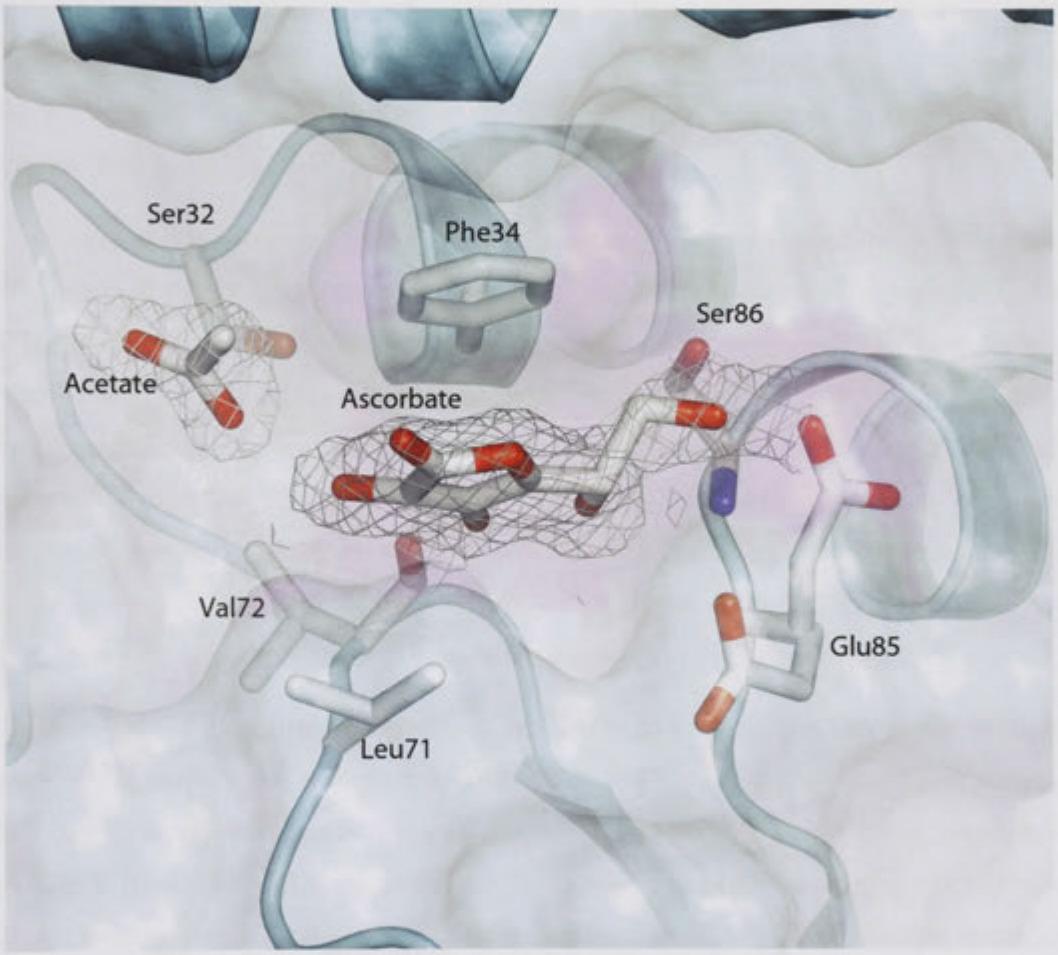


Figure 3.5
The observed binding mode of L-ascorbate in the active site of hGSTO1. The $2mF_O - DF_C$ electron density map corresponding to ascorbate and the adjacent acetate molecule is depicted as contoured isomesh at a level of 1σ. A translucent molecular surface, highlighted in magenta, demonstrates the shape of the binding pocket. The hGSTO1 fold is shown as a cartoon representation with relevant side chains as sticks, coloured according to element. The alternate conformers of E85 were refined with equal occupancy.

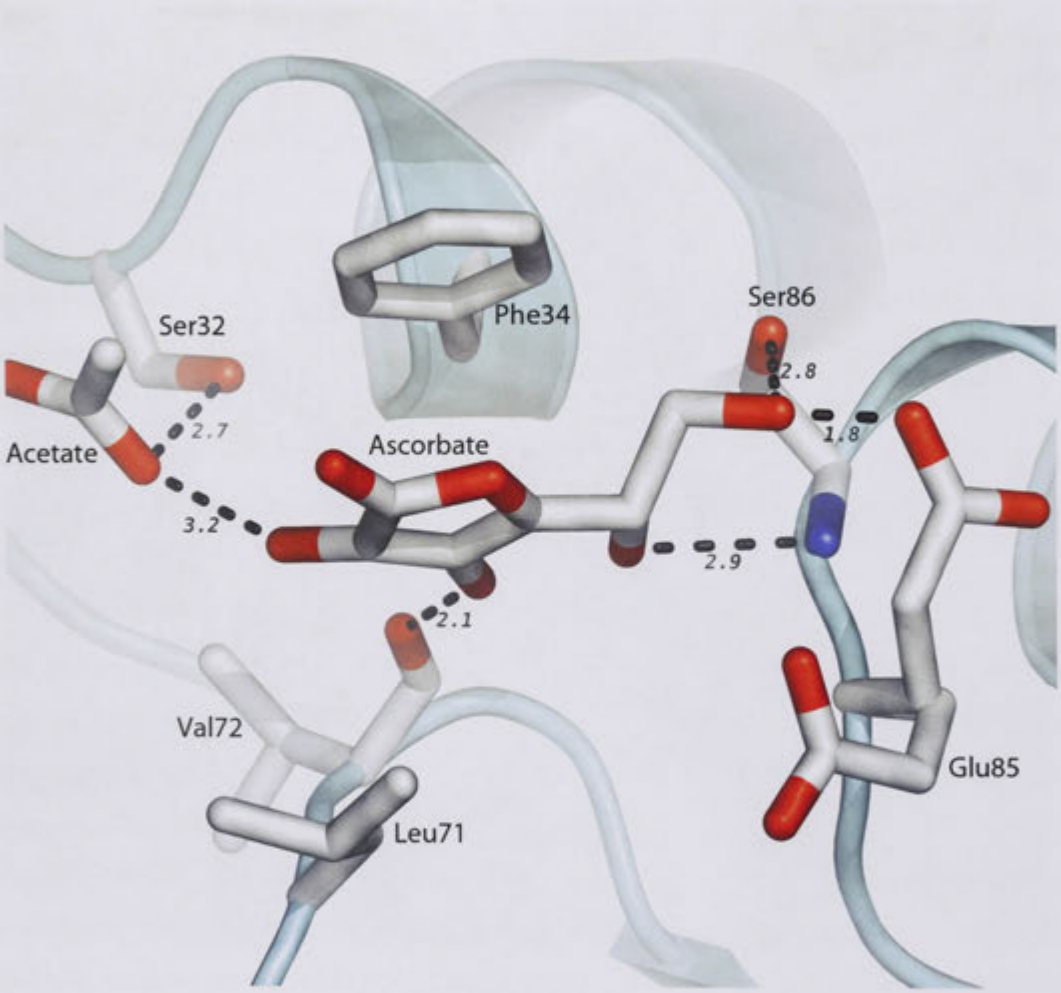


Figure 3.6
A view of L-ascorbate binding equivalent to Figure 3.5 on the preceding page, with the molecular surface removed and polar interactions shown as dashed lines. The crystallographically observed distances of each is given in Å. Solvent molecules have been omitted for clarity.

Indeed, such a hypothesis has precedence in the literature. Kinetic studies of the DHAR of spinach chloroplasts has indicated uncompetitive inhibition by 20 mM ascorbate with respect to dehydroascorbate. GSSG was also found to participate in product inhibition, as a competitive inhibitor with respect to dehydroascorbate, and a mixed type inhibitor with respect to GSH [54]. This plant DHAR shares 28% and 62% sequence identity and similarity with hGSTO1 respectively, and the two are likely to have evolved from a common ancestor. Both contain a conserved C-P-F-X-X-R active site motif (Figure 3.8 on page 56) and a search of the Conserved Domain Database (CCD) [55] revealed a shared GST_C domain (Figure 3.9 on page 57).

In light of this primary, and hypothesised secondary structural similarity, the catalytic mechanism of dehydroascorbate reduction in the two enzymes may also be conserved. The kinetic analysis of dehydroascorbate reduction by spinach chloroplast DHAR suggested a bi-uni-uni-uni-ping-pong mechanism analogous to that shown in Figure 3.1 on page 44: the initial binding of dehydroascorbate forming a cysteinyl-thioketal complex, followed by nucleophilic attack of GSH to yield a mixed disulfide and the reduced product. The active enzyme is then rejuvenated by another molecule of GSH, concurrently producing GSSG in the process. The kinetic analysis of product inhibition by GSSG mentioned above is consistent with this sequence of substrate binding and product release [54].

This mechanism is disputed however, by other authors studying the functional divergence of GSTs in plants, who discovered 3 novel DHAR enzymes in *Arabidopsis thaliana* [56], based on sequence similarity to hGSTO1. Using *hGSTO1* as the query sequence of a BLAST (Basic Local Alignment Search Tool) search identified a sequence in rice with DHAR reductase activity (GenBank^{TM} accession number AB037970). Using this sequence to search the genome of *A.thaliana* resulted in four related sequences, one of which was later discovered to be a pseudogene, showing 60-70% sequence similarity with the rice gene. Subsequent cloning, expression in *E.coli* and Ni-affinity chromatography purification resulted in enzymes showing high DHAR activity (2000-15 600 $nanokatals \cdot mg^{-1} protein$) and moderate thioltransferase activity when assayed with HED (2-hydroxyethyl disulfide) (15-131 $nanokatals \cdot mg^{-1} protein$). Like hGSTO1, the conserved active site cysteine was shown to be essential to both activities, and lacked the glutathione conjugation activity towards CDNB characteristic of most other GSTs [56]. In contrast, this study could not find evidence of product inhibition by ascorbate, although the assayed concentration (5 mM) was lower than that used in the study of spinach DHAR [54] (20 mM) or the struc-

tural findings presented here (100 mM). While the enzyme-GSH mixed disulfide of the reaction mechanism could be observed using ESI-TOF MS (electrospray ionisation time of flight mass spectrometry), the intermediacy of the hypothesised thioketal-enzyme covalent intermediate could not, and no evidence of this species could be found in other studies. This led the authors to propose a new mechanism based on single electron transfer (SET) (Figure 3.7 on the next page). Such a process would include the generation of SDAH. A known intermediate, its resonance stabilisation gives rise to the powerful anti-oxidant properties of ascorbic acid. Subsequent abstraction of hydrogen from the cysteinyl-thiol of GSH would be consistent with the well established ability of thiol compounds to act as hydride donors during radical reactions. The resulting thiyl radicals, GS$^{\bullet}$ and ES$^{\bullet}$, would then spontaneously react to yield the ESSG mixed disulfide, observed in the AtDHARs (*Arabidopsis thaliana* dehydroascorbate reductases) by ESI-TOF MS, and in the hGSTO1 crystal structure [4]. In support of this mechanism, the authors cited chemical studies on dione systems analogous to dehydroascorbate, suggesting initial attack of the cysteine (Figure 3.1 on page 44, step 1) would lead to a dead-end thioketal-enzyme complex [57]. In addition, the known SET reductants Cu(I) and Fe(II), were recently shown to be capable of reducing L-dehydroascorbic acid to L-ascorbic acid, and its c4 epimer, L-erythroascorbic acid [58]. Both of these species are also formed in the non-enzymatic reaction of dehydroascorbate with GSH, indicating that a stereospecific delivery of hydride to the enolic tautomer of the radical intermediate would feature as part of the catalytic mechanism (Figure 3.7 on the next page) [56]. If accurate, the high level of sequence conservation observed between plant DHARs and enzymes of disparate species such as hGSTO1 suggest it may be a conserved mechanism, evolved from an ancestral fold responsible for binding GSH. The high degree of sequence similarity between the divergent plant DHAR and GST enzymes with hGSTO1 is exemplified graphically in Figure 3.8 on page 56.

It should be noted that the BLAST search with the *hGSTO1* sequence by Dixon *et.al.*, [56] also discovered the proteins In2-1 in maize and Cla30 in wheat. These also bear sequence similarity to GSTs, contain a cysteine active site residue within a conserved GSH binding site and lack conjugating activity to CDNB. Subsequent BLAST searches with their respective gene sequences revealed that they are more similar to the mammalian Omega class than any plant GST, but were sufficiently different to justify classification within their own class, designated Λ (Lambda). Searching within the *A. thaliana* genome with the *In2-1* gene sequence revealed two sequences of a high homology, designated *gstl1* and *gstl2*. In

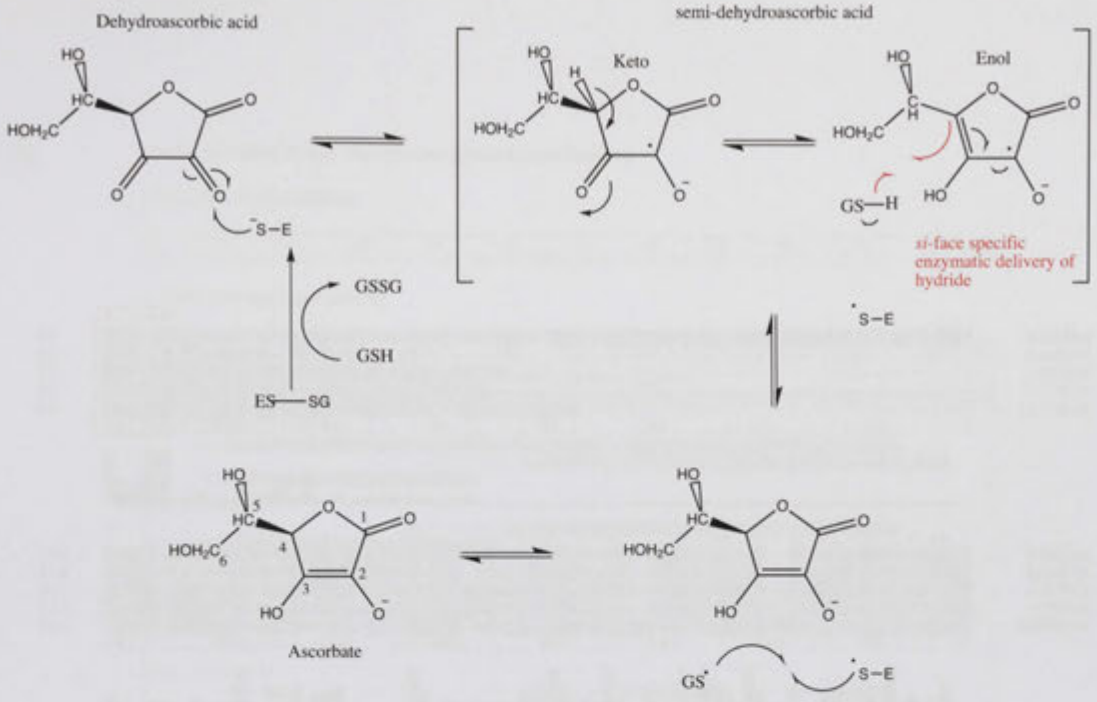


Figure 3.7
The SET mechanism of dehydroascorbate reduction as proposed by Dixon *et.al.*,[56]. A similar mechanism is observed non-enzymatically in which both c4 epimers of ascorbic acid are generated [58], implying that the enzymatic mechanism would feature the *si* face specific delivery of hydride to the enolic tautomer of the second intermediate [56].

contrast to the AtDHARs, cloning and purification of the resulting gene products resulted in polypeptides that displayed no dehydroascorbate activity, but retained thioltransferase activity as assayed with HED [56].

While the specific activity of hGSTO1-1 towards this substrate is approximately 3-fold lower than that observed for MMA^V, the dehydroascorbate activity of hGSTO2-2 has been found to be 70-100 fold higher. There is 64% sequence identity shared by the two enzymes [17], and by comparing the ascorbate bound structure of hGSTO1 with the apo-coordinates of hGSTO2 (Dr. Huina Zhou, unpublished data), it appears that all but one of the side chains associated with ascorbate binding are conserved. The lone exception is the replacement of leucine 71 with a histidine in hGSTO2, indicating an important contribution of this residue to the varying activity towards dehydroascorbate observed. Another feature of interest is the shift in helices α9-10 by several angstroms to provide a more 'open' active site within the hGSTO2 structure. The labile nature of this C-terminal extension was also observed when comparing the 3 monomers within the asymmetric unit of the hGSTO1Δ155

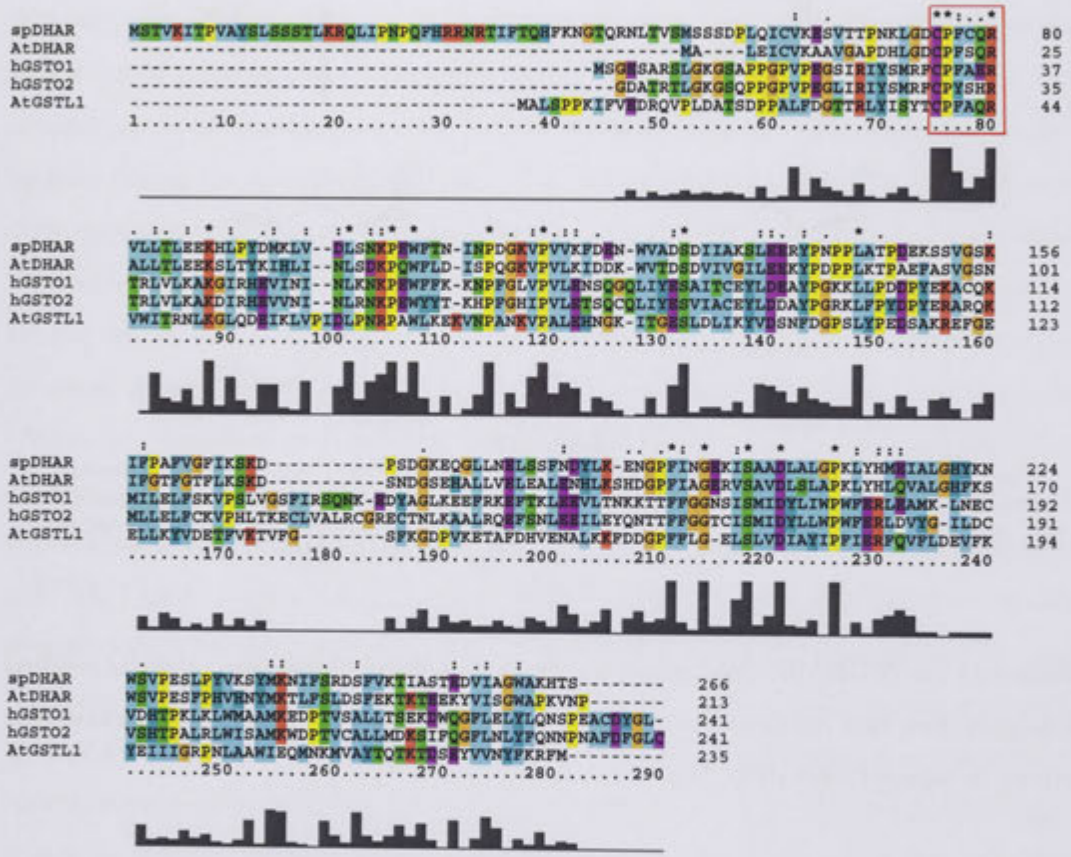


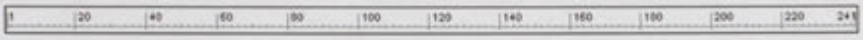
Figure 3.8

The sequence comparison of human GST omega enzymes with related DHAR [54, 59, 56] and GSTL [56] enzymes from plants discussed in the text. Amino acid residues are coloured according to chemical commonality and corresponding sequence similarity is indicated by the gray column graph. The highly conserved active site motif is highlighted by a red box. This highly conserved sequence similarity between plant and mammalian enzymes involved in dehydroascorbate reduction may be reflective of a structure and mechanism that has survived divergent evolution from an ancestral fold associated with GSH binding. This figure was generated using the program ClustalX [60].

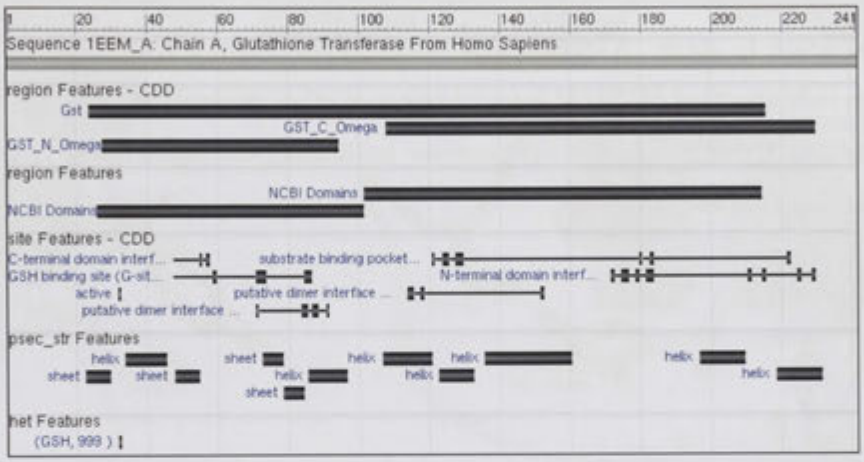
A

Chain A, Glutathione Transferase From Homo Sapiens

gi|10120947|pdb|1EEM|A



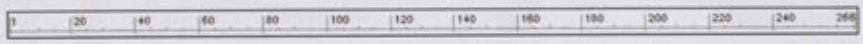
1 : 241 (241 residues shown)



B

dehydroascorbate reductase [Spinacia oleracea]

gi|10952512|gb|AAG24945.1|AF195783_1



1 : 266 (266 residues shown)

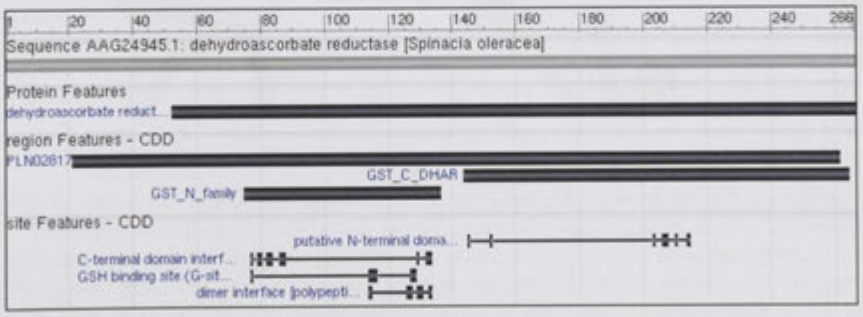


Figure 3.9

The observed sequence similarity between omega class GSTs and plant enzymes involved in dehydroascorbate reduction (Figure 3.8 on the facing page) is exemplified by commonalities within the CDD [55]. A comparison of conserved domains within hGSTO1 with the DHAR of spinach chloroplasts [54] (A and B respectively) demonstrates a conserved GST_C domain and GSH binding motif. The graphical representations of conserved structural domains above was generated by the graphics interface of the National Centre for Biotechnology Information (NCBI) proteins database (http://www.ncbi.nlm.nih.gov/guide/proteins/).

crystal structure (Appendix A on page 181), and may play an essential role in catalysis. Interestingly, this region contains a conserved tyrosine residue, at position 229 and 230 in hGSTO1 and hGSTO2 respectively, that could function as the yet unidentified hydrogen bond donor (Section 1.1.3 on page 8) responsible for stabilising the C32 or GSH associated thiolate during omega mediated catalysis (Figure 3.10 on the next page). While it's position and rotamer conformation within Omega structures to date have shown it to be outside the range of hydrogen bonding, the apparent mobility of the C-terminal domain may facilitate this interaction to occur. A survey of published mammalian dehydroascorbate reductase activities indicates that hGSTO2 is far in excess of other enzymes previously characterised [25], implying it may be the principal enzyme facilitating the regeneration of this essential antioxidant within the body.

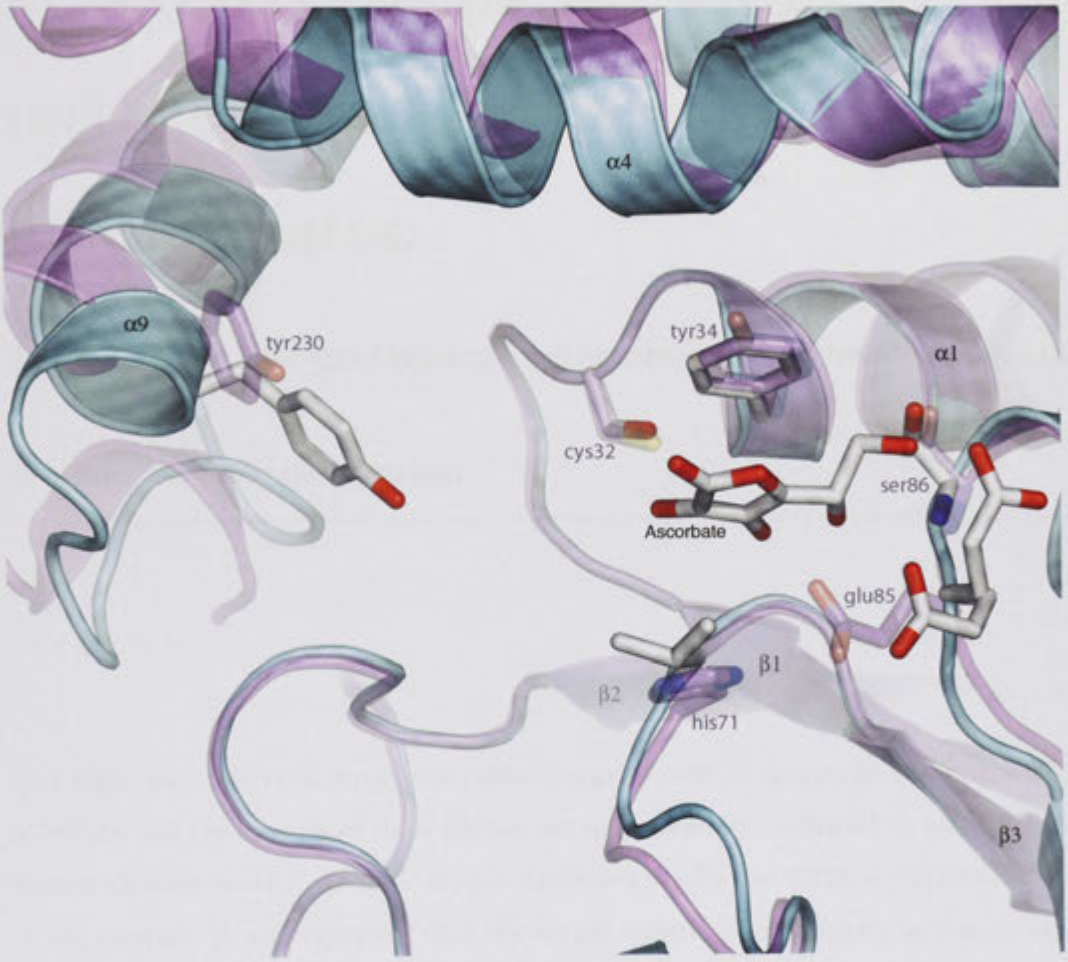


Figure 3.10

The hGSTO1 active site complexed with ascorbate (Figure 3.6 on page 52) compared to that of hGSTO2. The Structural fold of hGSTO1 is shown in a cyan cartoon with amino acid side chains expected to be relevant in dehydroascorbate reduction shown in stick representation, coloured according to atom type. The equivalent residues and fold of hGSTO2 is overlayed in magenta. An abstracted view of the active site allows visualisation of the conformational shift of the labile C-terminal extension comprised of α-helcies 9 and 10. Structural plasticity of this region was also implicated in the structural and folding studies of the hGSTO1Δ155 enzyme we have recently published (Appendix A on page 181). It is possible this movement may facilitate the stabilisation of an active site thiolate in the native enzyme via translocation of a conserved tyrosine residue (at positions 229 and 230 within hGSTO1 and hGSTO2 respectively) to within hydrogen bonding distance of the active site cysteine, or a molecule of GSH during catalysis. I am indebted to Dr. Huina Zhou for allowing me to use the coordinates of hGSTO2 prior to publication.

Chapter 4

hGSTO1-1 complexed with methylarsonate reduction intermediates:

Evidence for the glutathione intermediate pathway of human arsenic biotransformation

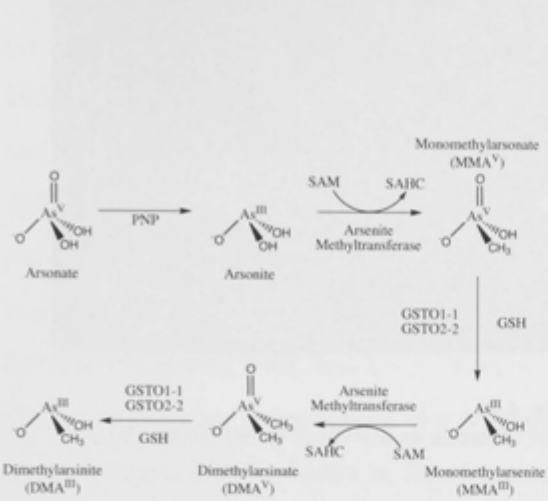
Two high quality crystal structures contain organic arsenic species in the active site of hGSTO1, and the identity of these species was unequivocally confirmed by anomalous difference electron density. As these crystal structures involve the C32A active site mutation of the enzyme, it was expected that the bound substrate, monomethylarsonate, would be observed relatively unmodified. Strikingly, the observed compounds appear to be the non-enzymatic products of methylarsonate reduction. The dithiol DTT, present in the experimental conditions, appears to have been sufficient to reduce the arsenic center to the +3 oxidation state, while the unaffected glutathione conjugating activity of the enzyme has allowed substitution of its hydroxyl ligands with molecules of GSH. These structures confirm that the hGSTO1 active site has a hight affinity for trivalent mono-, and possibly diglutathionyl methylarsenite, suggesting they would feature during it's catalysis. Implications to the understanding of human arsenic metabolism are discussed.

Coincidentally, one of the crystal structures has also revealed a derivative of the model substrate, CDNB, concurrently revealing several aspects of the active site dynamics within the 'H-site', implicated in its catalysis.

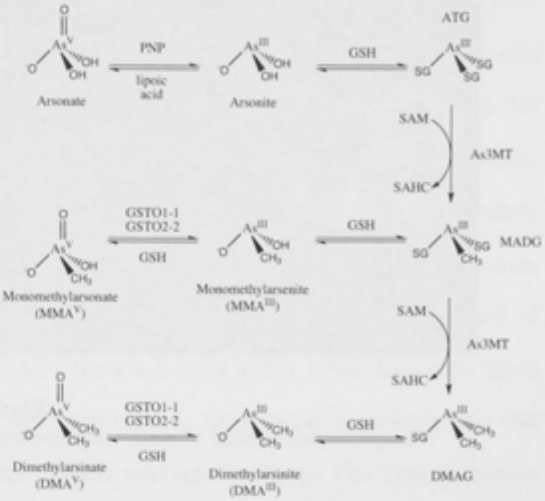
4.1 Introduction

Through the work of Frederick Challenger and his colleagues in the first half of the twentieth century [61], a conceptual model of the mammalian metabolism of inorganic arsenic (iAs) was developed that involved successive cycles of reduction, followed by oxidative methylation. Consistent with this hypothesis, Crecelius [62] was the first to conclusively observe methylated arsenicals in the urine of volunteers who ingested water or wine containing iAs^{III} or iAs^{V}. Contemporary research that observed these species in the urine of humans exposed to an environmentally high arsenic burden [63, 64, 65] was also consistent with this hypothesis. While the main metabolites were found to be the pentavalent species, monomethylarsonate (MMA^{V}) and dimethylarsinate (DMA^{V}), some authors also reported the presence of the trivalent monomethylarsenite (MMA^{III}) and dimethylarsinite (DMA^{III}) as well. While Purine Nucleoside Phosphorylase (PNP) purified from human liver was found to reduce $iAs^{V}$to iAs^{III} in the presence of the natural dithiol lipoic acid, it was unable to do so given organic varieties of pentavalent arsenic as a substrate [66]. The subsequent discovery that Omega GSTs were capable of catalysing this step in the proposed biotransformation lead to the proposal of the scheme shown below (Figure 4.1a on the facing page) [25].

This scheme has had to be reconsidered however, in light of recent findings that a common synthetic method used to create the organic trivalent arsenical standards of the HPLC (high pressure liquid chromatography) chromatographic coelution/ICP-MS (inductively charged plasma mass spectrometry) method of identification [67], results in a thiol-containing arsenical as the major product [68]. Indeed, it has become apparent that the MMA^{III} and DMA^{III} identified in urine, were likely the hydrolysis products of the corresponding complexes with GSH and that furthermore, the presence of methylated pentavalent arsenicals may have resulted from their spontaneous oxidation subsequent to sample collection [69]. Chromatographic experiments under conditions aimed to replicate those of a protein-free mammalian hepatocyte cytosol, indicates that trivalent arsenicals spontaneously form complexes with GSH and that the methylated varieties of these complexes are more stable than their inorganic counterparts [70]. It is now clear that trivalent, diglutathionylated arsenicals, are obligatory substrates of the enzyme responsible for arsenic methylation in humans, Arsenic 3+ methyltransferase (hAS3MT, also known as cyt19) [71, 72]. This has led to a new scheme of arsenic metabolism being proposed by Hayakawa and colleagues (Figure 4.1b on the next page) [71]. The results of this research suggest that



(a) The 'oxidative methylation' scheme of arsenic metabolism adapted from Schmuck *et al.*, [25].



(b) The 'non-oxidative methylation' scheme of arsenic biotransformation adapted from Hayakawa *et al.*, [71].

Figure 4.1
Two contemporary schemes describing the biotransformation of iAs within the human body. (PNP; Purine Nucleoside Phosphorylase, As3MT; Arsenic(III) Methyltransferase, SAM; S-adenosyl methionine, SAHC; S-adenosylhomocysteine, GSAHC; glutathionyl S-adenosylhomocysteine, ATG; arsenic(III) triglutathione, MADG; methyl arsenic(III) diglutathione, DMAG; dimethyl arsenic(III) glutathione).

of these two postulates put forth in the literature to describe iAs metabolism in humans, the latter is the most physiologically accurate.

The particular affinity of arsenic for thiol ligands was first identified in studies involving its toxic effects on nucleated red blood cells in the early 1900s, and evidence indicating that its target may be the thiol dependent urease [73]. This eventually lead to an antidote to Lewisite (chlorovinyldichloroarsine), a war gas that was developed in the period between the World Wars, but was never extensively deployed. Starting in 1939, British scientists used their fundamental understanding of As-thiol biochemistry to design dithiols capable of complexing this toxin and removing it from the body. The result was British Anti-Lewisite (2,3-dimercaptopropanol), an early attempt at directed drug design [73].

Chronic arsenic poisoning via contaminated drinking water is currently a global health pandemic [18]. At present, approximately 100 million people in India, Bangladesh and other developing nations are drinking water that contains iAs at levels up to 100 times the safe limit of 10 µg per liter advised by the World Health Organisation (WHO) [74]. The ingestion of arsenic occurring in these parts of the world have been referred to as the largest mass poisoning in history [75]. The resulting toxicology of this vast population includes the increased development of skin lesions, heart disease and several cancers, although little

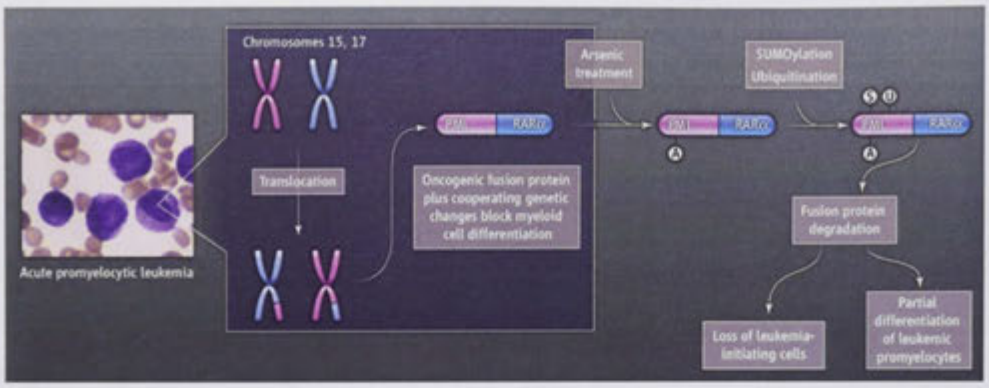


Figure 4.2
The pathomechanism of acute promyelocytic leukemia as described by SC Kogan [80].

is known about how these pathologies are affected [71, 76].

Paradoxically, the metabolism of arsenic has also become of considerable interest due to its recent use as a therapeutic. Although used in traditional Chinese medicine for thousands of years [77], arsenic trioxide (ATO) has recently been employed in the treatment of acute promyelocytic leukemia (APL) to dramatic effect (Figure 4.2). As the most malignant form of acute leukemia, its symptoms include a severe bleeding tendency and fatality often in a matter of weeks. In combination with *all-trans* retinoic acid (ATRA) however, this quite simple compound has allowed the percentage of complete remission to be raised beyond 90% [78]. The cause of APL is a translocation mutation between chromosomes 15 and 17, resulting in a fusion of the promyelocytic leukemia (*PML*) and retinoic acid receptor alpha ($RAR\alpha$) genes (Figure 4.2) [79]. While the native PML protein is considered a tumour suppressor, regulating the expression of genes involved in differentiation while compromised DNA is repaired, the resulting fusion product has quite the opposite effect. Joined to RARα, it's ability to block cellular differentiation at numerous places within the genome becomes constitutively turned on and the differentiation from promyelocytes to mature neutrophils is blocked. This results in the unchecked replication of these immature white blood cells and subsequent leukemia [80]. Although previous research has shown that the cytotoxic effect of ATO may be associated with the generation of reactive oxygen species (ROS) [81][82], conclusive studies have recently demonstrated that the efficacy of ATO is principally via direct binding to cysteine residues within the 'zinc-finger' DNA-binding region of PML. Combined with ATRA binding to RARα, this is thought to cause a conformational change that favours interaction with small ubiquitin-like protein modifiers (SUMOs), marking the polypeptide for degradation [83][84].

PMLs role in regulating gene expression has also been implicated in other forms of cancer. Chronic myeloid leukemia (CML) for instance, exemplifies the hypothesis that some leukemias are sustained by a core population of cells that posses stem cell like character and are resistant to conventional therapy. Evidence of these 'Leukemia Initiating Cells' (LICs) continues to grow, and it has been suggested that they are at the source of many relapses. After conventional therapy has appeared to eradicate the cancer cell population, it is thought that this quiescent population again propagates the disease after a period of latency. Targeting PML has been shown to eradicate such cells [85]. PML has also been shown to be a key regulator of the p53 tumour suppressor, mutations in which are the most common genetic factor in human cancers. PML actively increase the transcription of mutant p53 and is important in the development of the oncogenic 'gain of function' properties it often displays [86]. This raises the possibility that this highly effective treatment of APL, in combination with the current treatment of these forms of cancer, may reduce their malignancy and minimise the probability of relapse. A principle objective of this research was to resolve the chemical species of arsenic resulting from hGSTO1-1 catalysis, since it may be these and/or other metabolic derivatives of ATO that affect its anti-cancer properties via interaction with PML. A synthetic glutathione derivative of trivalent arsenic, *para*-4-[N-(S-glutathionylacetyl)amino]phenylarsenoxide, displays potent cytotoxic anti-cancer activity and is currently undergoing clinical trials [87].

Soaking of native hGSTO1 crystals with high concentrations of MMA^V and DMA^V consistently resulted in an inability to observe the arsenicals bound in the hGSTO1 active site. This result was independent of the omission or inclusion of glutathione at varying concentrations, and similarly unresponsive to combinations of the methylated arsenicals with the substrate analogue, glutathione sulfonic acid. Similarly, co-crystallisation experiments with these arsenicals in the presence or absence of glutathione, or its inactive analogue, were also unsuccessful. Analogous experiments with the C32S mutant were also unproductive. The reasons for this are unclear, but will be discussed in more detail below (Section 4.3 on page 71). Soaking and co-crystallisation experiments with the C32A mutant however, did achieve a crystallographic observation of arsenic in the hGSTO1 active site, and was unequivocally confirmed using anomalous difference density (Section 2.5.4 on page 33). Despite the enzymatic inactivation resulting from the mutation, the arsenic centre was surprisingly discovered with an altered oxidation state and ligand substitution. Here, two high quality crystal structures are presented. The first confirms the specific

binding of glutathionyl methylarsenite in the hGSTO1 active site, while the second suggests a high affinity for diglutathionyl methylarsenite as well. Through a serendipitous combination of experiments on a single crystal, the former has also revealed the binding of an entirely different compound: several angstroms away from glutathionyl methylarsenite, residing in the hitherto putative H-site, a derivative of the model compound CDNB was observed, 1-hydroxy-2,4-dinitrobenzene (HDNB).

4.2 Experimental Methods

The alanine active site mutant had previously been made by my colleague Majorie Coggan [17], but first needed to be transferred to the expression vector pHUE [37] (Figure 2.2 on page 19 and Section 2.2 on page 18). Subsequent to expression in *E. coli* and purification by Ni-affinity chromatography, the binding of MMA^V to the hGSTO1 active site mutant was investigated via co-crystallisation and soaking experiments involving various concentrations of MMA^V, GSH and (tris(2-carboxyethyl)phosphine) (TCEP). In a particular co-crystallisation experiment, a crystal grown in the presence of MMA^V was subsequently 'back soaked' in a solution containing CDNB and GSH, yielding unexpected results.

4.2.1 Cloning of hGSTO1 C32A into the pHUE vector.

The alanine active site mutant was obtained in the pQE30 plasmid (Qiagen) and had to be cloned into the pHUE expression vector before protein could be purified for crystallisation. This was achieved via a complementary *XmaI*/*SacI* (NEB) mediated restriction digest of both. The digestion products were then separated on 1% agarose gels in electrophoresis buffer, both containing 0.5 μg/ml of ethidium bromide. The gels were subsequently visualized with a UV transilluminator at 366 nm. While the digested pHUE vector was easily discernible from a digest of DNA equivalent to 1 mL of overnight culture, digested DNA equivalent to 10 mL of overnight culture was needed to observe the *hGSTO1* cysteine 32 to alanine insert fragment. Both were found at the expected positions relative to the simultaneously loaded 'GeneRuler' DNA marker (Fermentas), visualized at 5.9 Kb and 0.7 Kb, respectively. After careful removal with a scalpel, the DNA fragments were purified using a 'QIAquick gel extraction kit' (Qiagen). Ligation with T4 DNA ligase (Roche) was then performed, followed by transformation into heat shock competent *E. coli* cells of strain XL1BL. These were immediately used to inoculate a LB agar plate (0.2 mg/ml Ampicillin) that was subsequently incubated overnight at 37°C. The resulting colonies were

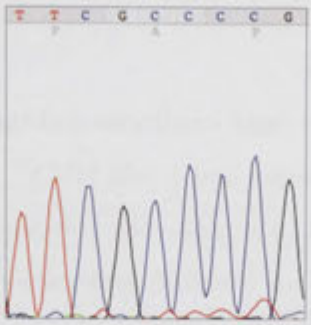


Figure 4.3
The DNA sequencing chromatograms of the C32A active site mutation, unequivocally confirming the TGC (cysteine) to GCC (alanine) nucleotide permutations.

used to make LB cultures for a 'Miniprep' kit (Qiagen) DNA extraction. A *SacII*/*NotI* restriction digest was subsequently performed and the products separated and visualized using an ethidium bromide agarose gel as above. As these two restriction sites occur exclusively within the *hGSTO1* gene and not within the pHUE vector, the identification of two fragments of DNA could confirm a correctly ligated gene construct, and not just an empty pHUE vector conferring resistance to the antibiotic. Indeed, the vector was observed between the 2 Kb and 3 Kb markers, and the insert observed parallel to the 0.8 Kb marker, as expected. A DNA sequencing reaction was then performed and submitted to the Biomolecular Resource Facility (BRF), JCSMR, for interpretation. The resulting chromatogram unequivocally confirmed the presence of desired TGC (cysteine) to GCC (alanine) nucleotide change (Figure 4.3).

4.2.2 Purification, crystallisation, data collection and refinement

Protein was purified to homogeneity as described in Chapter 3 (Section 3.2.2 on page 46). Crystallisation of the protein was then performed using the hanging drop/vapour diffusion method. Crystallisation conditions were optimised by screening against pH and precipitant concentration, based upon previously published conditions [4]. There are two crystal structures of hGSTO1-C32A involving species of methylated arsenic discussed in this chapter. The first was co-crystallised with 100 mM MMA^V pH 7.5, before being 'back soaked' in a solution prepared by combining 1 μL of 10 mM CDNB, 10 mM GSH, pH 7.5 and 1 μL well solution, and allowing it to equilibrate within the sealed reservoir of the hanging drop experiment for 48 hours. This step was due to previously solved structures under analogous

conditions failing to show binding of the arsenical, and the need at the time for crystals of hGSTO1-C32A to investigate the increased activity of this active site mutant to CDNB (Section 1.1.3 on page 8) [17]. This resulted in species derived from both experiments being observed in the active site.

The second was grown under 'apo' conditions and then transferred to an equilibrated soaking solution of reservoir solution and 5 mM MMA^V, 10 mM GSH and 5 mM TCEP, pH 7.5. The latter was included to simulate the reducing power of the wild type enzyme. Both datasets were collected at the Stanford Synchrotron Radiation Lightsource (SSRL), California, USA, using radiation corresponding to the anomalous absorption edge of arsenic (12680 eV). Initial phase estimates were calculated from the previously solved structure (PDB identifier: 1EEM) [4]. The details of the crystallisation, data collection and refinement of each are tabulated in Table 4.1 on the next page, and Table 4.2 on page 70, respectively.

Crystallisation	Experimental Parameters of hGSTO1-C32A, glutathionyl methylarsenite, HDNB complex.
Reservoir solution	2.2 M $(NH4)_2SO_4$
	100 mM Sodium Acetate pH 4.25
Crystallisation drop	1 µL hGSTO1-C32A (30 mg/ml)
	1 µL 100mM MMA^V pH 7.5
	1 µL reservoir
Soaking solution$^{\mp}$	1 µL 10mM CDNB pH 7.5
	1 µL 10mM GSH pH 7.5
	1 µL reservoir
Cryoprotectant solution	2.2 M Li_2SO_4
	100 mM Sodium Acetate pH 4.25
	Sequential transfer from 5%—->10% Glycerol solution
Data collection	
Resolution limits, Å	28.08-1.47 (1.52-1.47†)
Space group	$P3_121$
Unit Cell, Å	a=b=56.17 c=139.54
Total reflections	562 469
Unique reflections	44 388
Completeness, %	99.8 (98.2†)
I/σ	47.5 (3.0†)
Redundancy	12.7 (5.6†)
Refinement	
R_{cryst}/R_{free}, %	12.9/15.9
rmsd bonds, Å	0.012
rmsd angles, °	1.389
Number of waters	269
Ligands	SO_4^{2-}, glycerol, acetate, glutathionyl methylarsenite, HDNB, Cl^-

†, the highest resolution shell. $^{\mp}$, allowed to equilibrate for 48 hours.

Table 4.1
Crystallisation details and structure determination statistics of the hGSTO1-C32A, glutathionyl methylarsenite, HDNB complex

Crystallisation	Experimental Parameters of hGSTO1-C32A, diglutathionyl methylarsenite complex.
Reservoir solution	2.2 M $(NH4)_2SO_4$
	100 mM Sodium Acetate pH 4.25
Crystallisation drop	2 µl hGSTO1-C32A
	2 µl reservoir
Soaking solution$^{\mp}$	1 µl 5mM MMA^V, 5mM TCEP, 10mM GSH, pH 7.5
	1 µl reservoir
Cryoprotectant solution	2.2 M Li_2SO_4
	100 mM Sodium Acetate pH 4.25
	Sequential transfer from 5%—>10% Glycerol solution
Data collection	
Resolution limits, Å	50.00-1.46 (1.51-1.46†)
Space group	$P3_121$
Unit Cell, Å	a=b=57.08 c=140.1
Total reflections	499 079
Unique reflections	81 805
Completeness, %	92.5 (52.9†)
I/σ	14.5 (1.86†)
Redundancy	6.1 (2.9†)
Refinement	
R_{cryst}/R_{free}, %	17.4/22.2
rmsd bonds, Å	0.007
rmsd angles, °	1.074
Number of waters	115
Ligands	$SO_4{}^{2-}$, glycerol, acetate, diglutathionyl methylarsenite

†, the highest resolution shell. $^{\mp}$, allowed to equilibrate for 48 hours.

Table 4.2
Crystallisation details and structure determination statistics of the hGSTO1-C32A, diglutathionyl methylarsenite complex

4.3 Results and discussion

Serendipitously, the clearest, and most unequivocal crystal structure obtained to date of the interaction of hGSTO1 with a methylated arsenic species also revealed a derivative of the model substrate CDNB. Crystals of the hGSTO1-C32A active site mutant, previously co-crystallised with MMA^V, were subsequently soaked within a drop containing CDNB, well solution and glutathione overnight (Section 4.2 on page 66). The resulting crystal structure showed species originating from both experiments bound within the active site of hGSTO1 (Figure 4.4 on page 73). Despite the active site mutation, the arsenic species was observed reduced to the +3 oxidation state, with one of its hydroxyl ligands displaced by a molecule of glutathione. Hypothesised to have been displaced from a molecule of 1-glutathionyl-2,4-dinitrobenzene (Figure 4.14 on page 86), the resulting compound, HDNB, was observed binding within the previously hypothesised 'H-site' of the enzyme (Figure 1.7 on page 11). Its Cl^- leaving group was observed bonded to the side chain of an arginine residue 10 Å away (Figure 4.8 on page 77) [4].

While the structure referred to above shows a single glutathione complexed to a trivalent arsenic atom, other results support the suggestion of Hayakawa [71], that the metabolism of monomethylated arsenic occurs via a diglutathionylated trivalent species (Figure 4.1b on page 63). Although the high resolution of the crystal structure allowed the coordination by two As-S bonds to be identified with certainty, the presence of the dithiol DTT in the experimental conditions, combined with ambiguous electron density, allowed only one of these to be assigned to glutathione unequivocally.

4.3.1 Structural features of glutathionyl methylarsenite and HDNB/Cl^- binding in the active site of hGSTO1

Glutathionyl methylarsenite was observed with an occupancy of 0.58 with native glutathione occupying the remaining 48% of the crystal lattice (Figure 4.5 on page 74). Using X-rays at the anomalous edge of arsenic absorption, coupled with the high resolution of diffraction data obtained, allowed the former species to be identified absolutely (Figure 4.2 on page 66 and Figure 4.6 on page 75). Although binding of the glutathione moiety in both instances was as previously observed [4], the crystal structure reveals a novel polar interaction between Y229 and the lone electron pair of arsenic, mediated by a bridging water molecule (Figure 4.6 on page 75).

Only a few angstroms away, a derivative of the model substrate CDNB was observed

within the predicted 'H-site' [4]. A lack of strong $2mF_O - DF_C$ electron density corresponding to the chlorine atom was indicative of substituion to a hydroxy group, to form HDNB, and refined to an occupancy value of 0.75. Indeed, the displaced chloride ion (occupancy = 0.71) was observed binding to the guanidinium group of R37, approximately 10 Å away. Modeling a water molecule in this position consistently produces a large positive peak in the $mF_O - DF_C$ electron density and an abnormally low B-factor. The interdependence of these species is attested to by the close correlation of their independently refined occupancy values.

In addition, the binding of HDNB is associated with a conformational change in two neighbouring side chains. R132 has moved several angstroms to form a polar interaction with the compounds hydroxyl group, at the same time forming a salt bridge interaction with a novel sulfate ion. Similarly, W222 has adopted an alternate rotamer conformation such that its indole nitrogen is now orientated away from the binding pocket to form a hydrogen bond with the sulfur of an alternate conformation of M187, effectively demonstrating a mechanism that may facilitate Omega class binding promiscuity (Figure 4.8 on page 77).

Interestingly, this latter interaction with M187 has been facilitated by an intriguing conformational change within the α-helix 6 to 7 junction, located immediately behind the H-site. The α-helical character of this region has been abolished, causing residues 183-196 to adopt significantly different conformations to those previously observed in the complex with GSH (Figure 4.9 on page 78) [4]. Although less pronounced, this conformational change is also induced by binding of GSSG (Figure 5.5 on page 108).

4.3.2 Hypothesised binding of diglutathione methylarsenite

Although the electron density is not definitive, another high resolution structure suggests the glutathionyl methylarsenite observed above may subsequently progress to a diglutathionyl species within the hGSTO1 active site. This is the chemical form suggested by Hayakawa and Thomas *et. al.*, [72, 71], to be necessary for hAS3MT catalysis. The 1.5 Å structure has been modeled with 2 molecules of native glutathione in alternate conformations, and one molecule of diglutathionyl methylarsenite overlaid in equivalent positions within the active site. The three molecules have been refined to occupy approximately one third of the crystal lattice each. For clarity, this is represented with colour coding in Figure 4.10 on page 80. While one glutathionyl ligand of diglutathionyl methylarsenite is seen to occupy it's previously observed position in the G-site, the second has been modeled

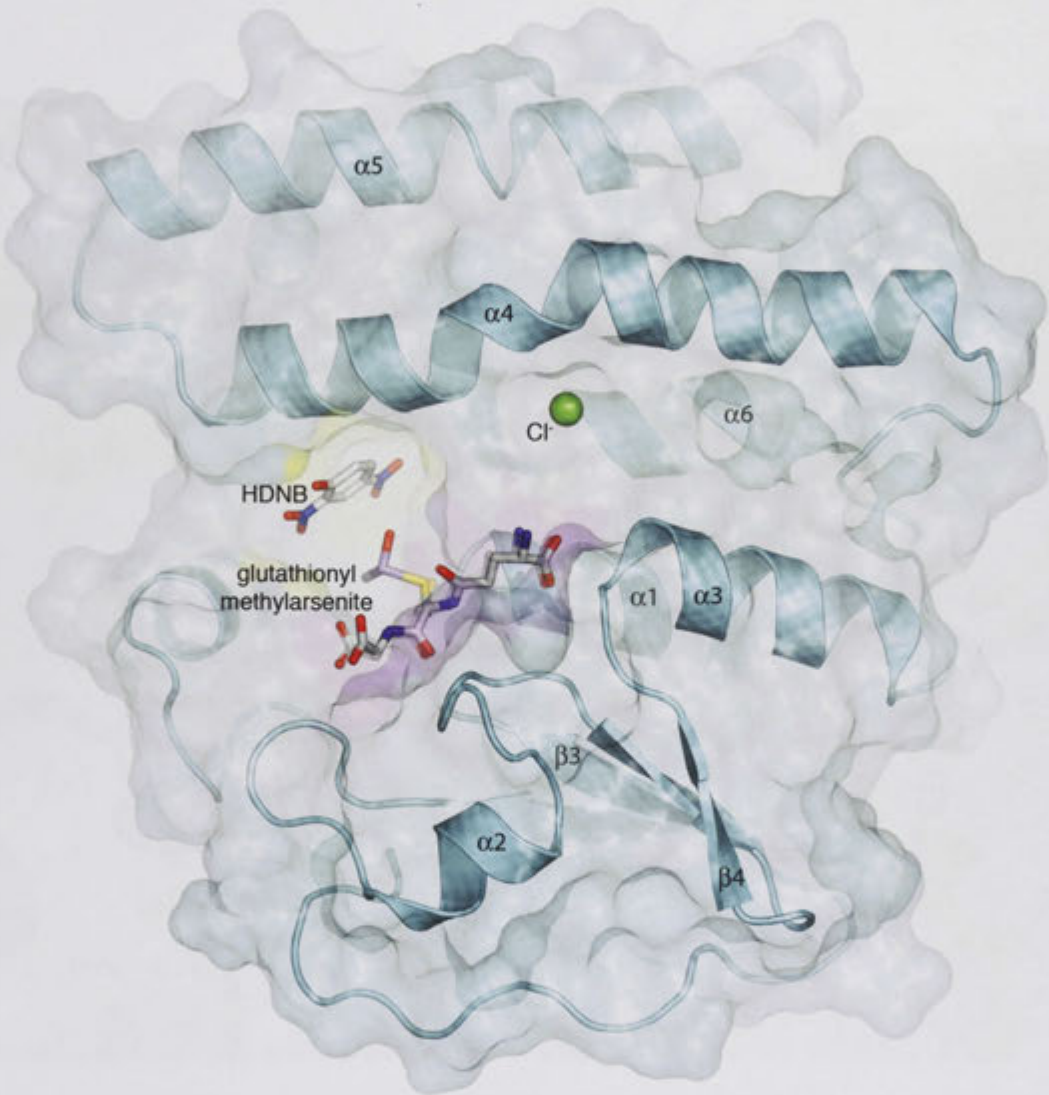


Figure 4.4

The 1.47 Å resolution crystal structure of hGSTO1 in complex with glutathionyl methylarsenite, glutathione and a derivative of CDNB, HDNB. All three ligands are shown in stick representation and were observed at partial occupancies, refining to occupancy values of 0.58, 0.42 and 0.75 respectively. The disassociated chloride ion (green, occupancy = 0.71) was observed 10 Å away, forming a salt bridge interaction with an arginine guanidinium group. Due to the crystal used for the CDNB/glutathione soaking experiment previously being co-crystallised with MMA^V, species from both experiments were observed in the resulting crystal structure. The crystallographic fold of hGSTO1 is shown as a cyan cartoon overlaid with a translucent molecular surface. The binding of glutathionyl methylarsenite and HDNB are highlighted with magenta and yellow respectively. Both binding sites are within Å of each other and occur within the crevice formed at the C-terminal (upper hemisphere) and N-terminal (lower hemisphere) domain interface.

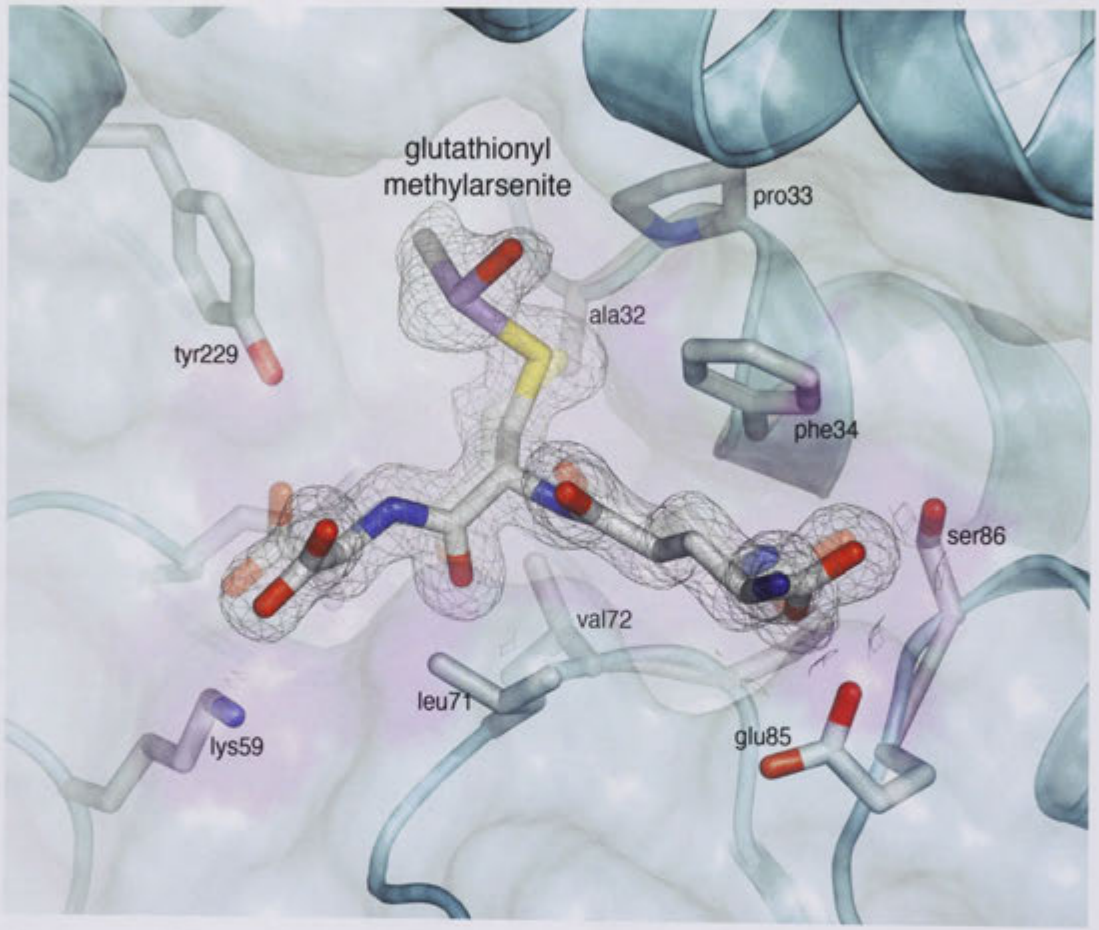


Figure 4.5

Glutathionyl methylarsenite and associated side chains are shown in stick representation. The binding of native glutathione is shown as a translucent superposition. The arsenic derivative was refined to an occupancy value of 0.58, with native glutathione occupying the remaining 42% of the crystal lattice. The weighted $2mF_O - DF_C$ electron density corresponding to the former is depicted as gray isomesh, contoured at a level of 1σ. hGSTO1 is shown as a cyan cartoon. A translucent molecular surface highlights the binding pocket of glutathionyl methylarsenite in magenta.

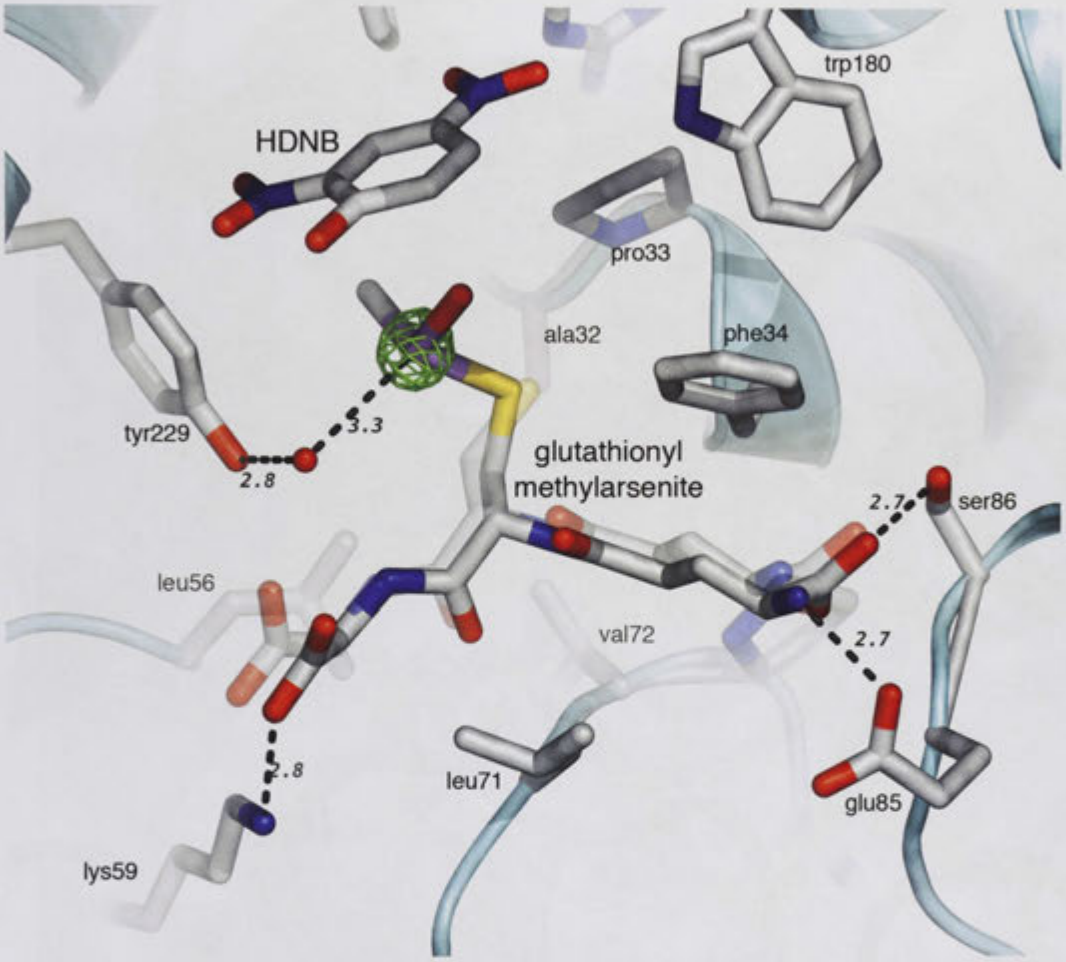


Figure 4.6

A view of the polar interactions associated with glutathionyl methylarsenite binding in relation to HDNB. The crystallographic distance of each is given in Å. A solvent molecule mediating a hydrogen bonding interaction between the Y229 hydroxyl group and the arsenic lone pair is shown as a red sphere. Other solvent molecules have been omitted for clarity. The glutathione adduct of the trivalent methyl arsenical and associated side chain residues are shown in stick representation, coloured according to atom type. Shown in green isomesh contoured at 20σ, the anomalous difference density generated by the arsenic atom of glutathionyl methylarsenite confirms its identity unequivocally.

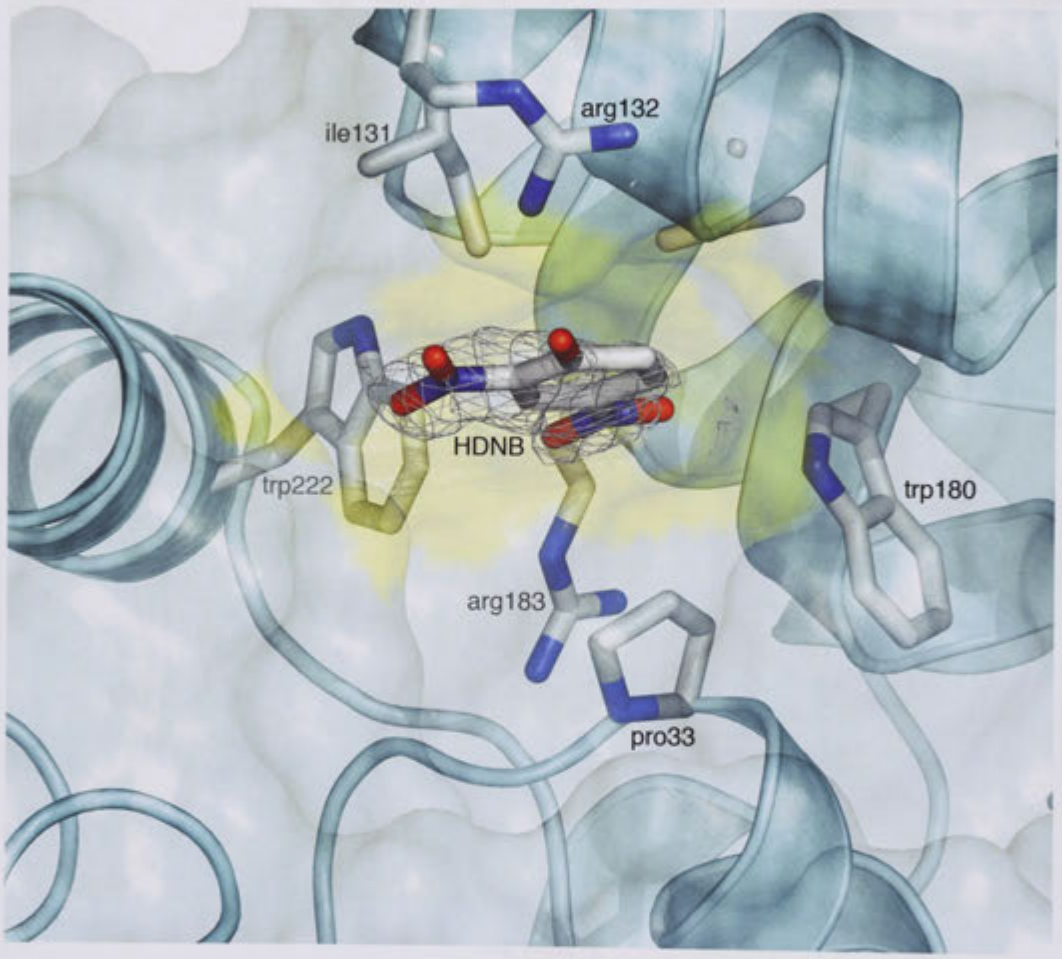


Figure 4.7

The binding of HDNB is shown in stick representation. The corresponding $2mF_O - DF_C$ electron density is depicted as gray isomesh, contoured at a level of 1σ. The crystallographic fold of hGSTO1 is shown as a cyan cartoon. A translucent molecular surface highlights the binding pocket of HDNB in yellow.

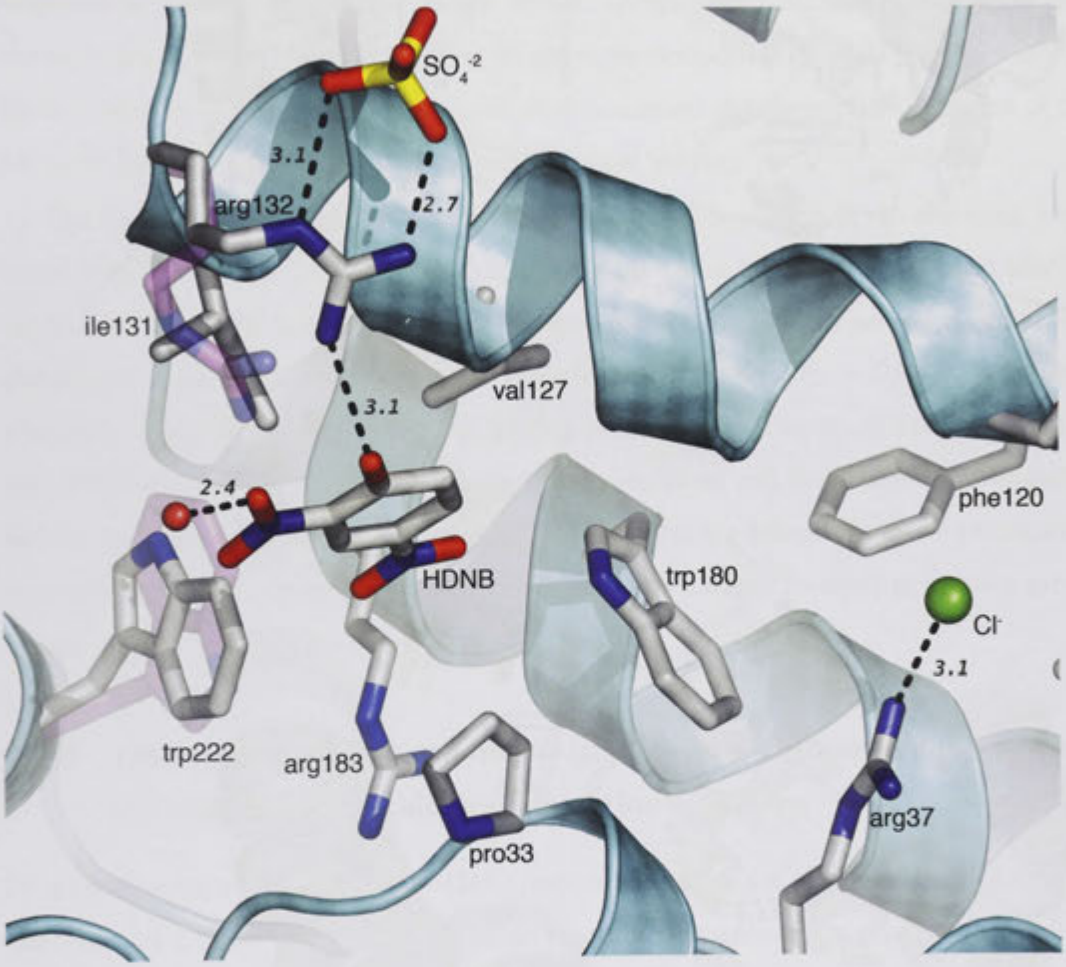


Figure 4.8

Details of the interactions between HDNB and the dissociated Cl^- ion with associated residues of hGSTO1 are shown in stick representation. Binding of HDNB has caused a dramatic conformational change in the side chains of R132 and W222. Their previously observed positions are overlaid in translucent magenta [4]. The former has been reoriented by several Å to form a polar interaction with the 1-hydroxyl group, and is being concurrently stabilised by the novel binding of a sulfate ion. The latter accommodates the hydrophobicity of the compounds aromatic core via adoption of an alternate rotamer conformation. This has reoriented its polar indole group away from the binding pocket, to form a hydrogen bonding interaction with the sulfur of a neighbouring methionine (Figure 4.9 on the next page). The *ortho*- and *para*-substituted nitro groups display limited polar interactions, the only identifiable hydrogen bond being with an ordered molecule of solvent (red sphere). The disassociated Cl^- is observed approximately 10 Å away, participating in a salt bridge with the guanidinium group of R37. Other solvent molecules have been omitted for clarity. Crystallographically observed distances are given in Å.

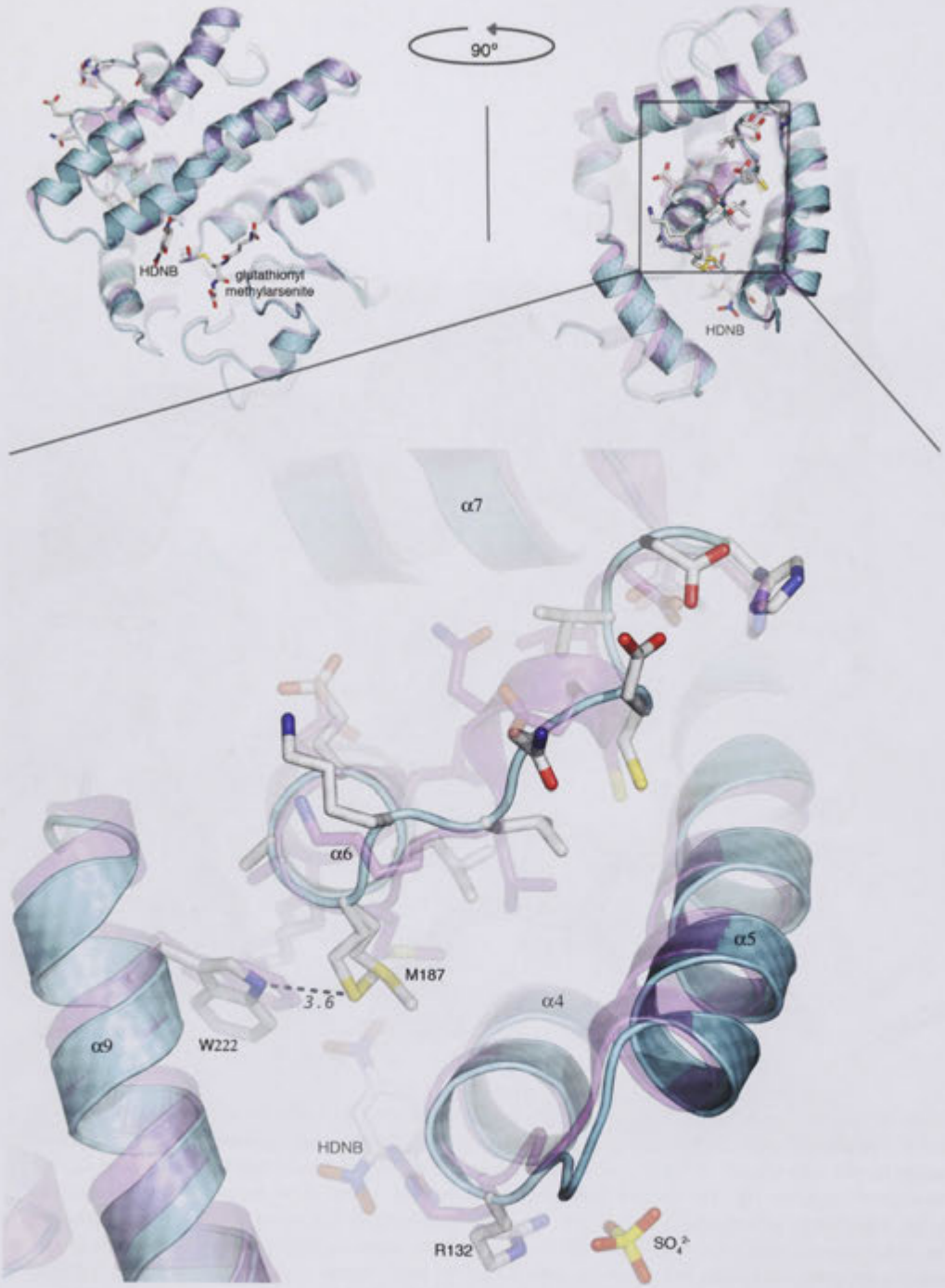


Figure 4.9

The α-helical character of the α6-α7 junction, posterior to the H-site, has been abolished by a ligand associated conformational change. The crystallographic fold of the hGSTO1 C32A methyl glutathionylarsenite/HDNB ternary complex (cyan) is compared with the aligned coordinates of the glutathione complex (magenta) [4]. HDNB and the associated H-site residues, W222 and R132, are shown in stick representation coloured according to atom type, as are residues 183-196. The weak polar interaction that is now formed between the reoriented indole group of W222 and the sulfur lone pair of an M187 alternate conformation is shown with a dashed line, the crystallographically observed distance is quoted in Å. Above, an abstracted view of the hGSTO1 structural fold shows the location of this region relative to ligand binding from orthogonal perspectives. This structural riposte to ligand binding was also observed in the 4NPG/GSSG ternary complex, but in a different conformation to the one observed here (Figure 5.5 on page 108).

to extend upwards towards the C-terminal domain. Here, its γ-glutamyl carboxylate may accept a hydrogen bond from the indole amine of W180 (Figure 4.11 on page 81), while the polar interactions of the glycinyl portion are hypothesised to be entirely mediated by surrounding solvent. In contrast to the ternary complex with HDNB mentioned above, the side chain of W222 has not adopted an alternate conformation, while the side chain of R132 is seen to be disordered. Associated conformational change within the region of the α6 to α7 helix junction is relatively minor (data not shown).

The lack of defined protein-ligand interactions with this portion of the arsenical, combined with its many rotatable bonds, may contribute to an inherently disordered binding mode and hence, the lack of conclusive electron density. While this second coordinating glutathione does not produce associated negative peaks in the $mF_O - DF_C$ map or unrealistically high B-factors in the atoms lacking $2mF_O - DF_C$ density, it should be noted that DTT was present within the experimental conditions and could also be responsible for the As-S density observed. Altering the experimental conditions of protein purification to exclude this non-physiological reducing agent subsequently resulted in a space group change upon crystallisation (Chapter 6 on page 131).

4.3.3 Discussion of the mechanistic implications for methylarsonate reduction and glutathione conjugation reactions

Proposed reaction mechanisms of MMA^V reduction *in vivo* are shown in Figures 4.12 on page 83 and 4.13 on page 84. While these reaction mechanisms are given for the native enzyme, the species highlighted in blue and red were modeled in crystal structures of the C32A active site mutant. Given that these chemical species were not present in the crystallisation conditions (Table 4.1 on page 69 and Table 4.2 on page 70) and that this active site mutation has been shown to abolish the characteristic thioltransferase activity of hGSTO1-1 [15, 17], the question that must be asked is how these compounds were formed at all? The foremost explanation is perhaps that their formation was non-enzymatic, and high affinity to the hGSTO1 active site subsequently allowed them to be observed crystallographically. Indeed, Percy [70] has confirmed the spontaneous formation and stability of trivalent glutathionyl methyl arsenicals under physiological conditions. There is the possibility however, that even without a functional active site residue, hGSTO1-1 may still have catalysed the conjugation of glutathione to the arsenic center separately to its reduction. As noted previously (Section 1.1.3 on page 8), the same mutation actually increases its

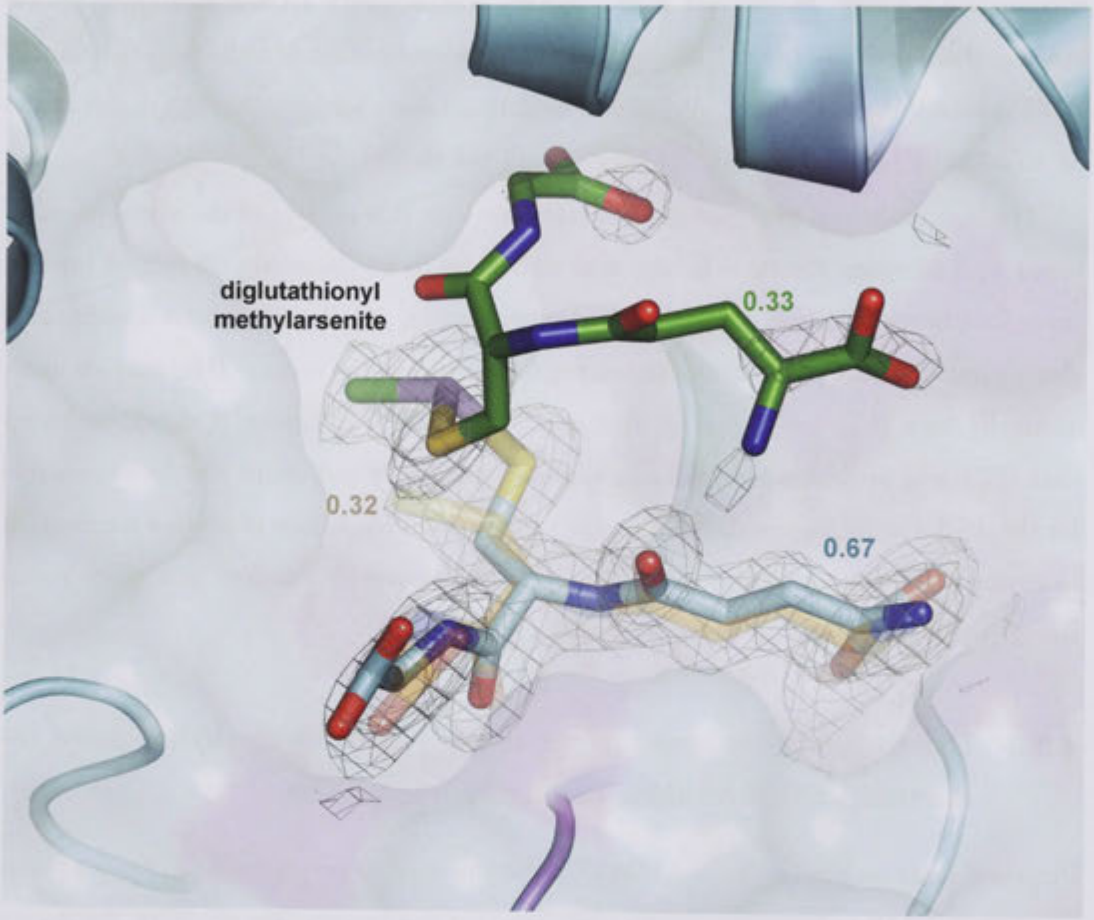


Figure 4.10

Shown in stick representation, the carbon atoms of the various glutathione molecules have been colour coded to allow their occupancy values to be distinguished. Within the 'G-site', unmodified glutathione is overlaid in orange (0.32 occupancy). Another molecule of glutathione with an alternate conformation of its sulfhydryl occupies the same position within the remainder of the crystal lattice (cyan), however, it has been modeled as part of a diglutathionyl methylarsenite molecule 33% of the time (green). Atoms are otherwise coloured according to atom type, with arsenic shown in magenta. The experimentally observed $2mF_O - DF_C$ electron density is shown as gray isomesh contoured at 1σ. While the density corresponding to most of the green glutathione molecule is not complete, the trivalent arsenic atom coordinated by two sulfur atoms is unmistakable. The high resolution of this crystal structure allows these characteristically long As-S bonds to be modeled with certainty, although there is the possibility that the latter could be due to coordination by a disordered molecule of DTT, also present in the experimental conditions. hGSTO1 is shown as a cyan cartoon overlaid with a translucent molecular surface. The binding site of diglutathionyl methylarsenite is highlighted in magenta.

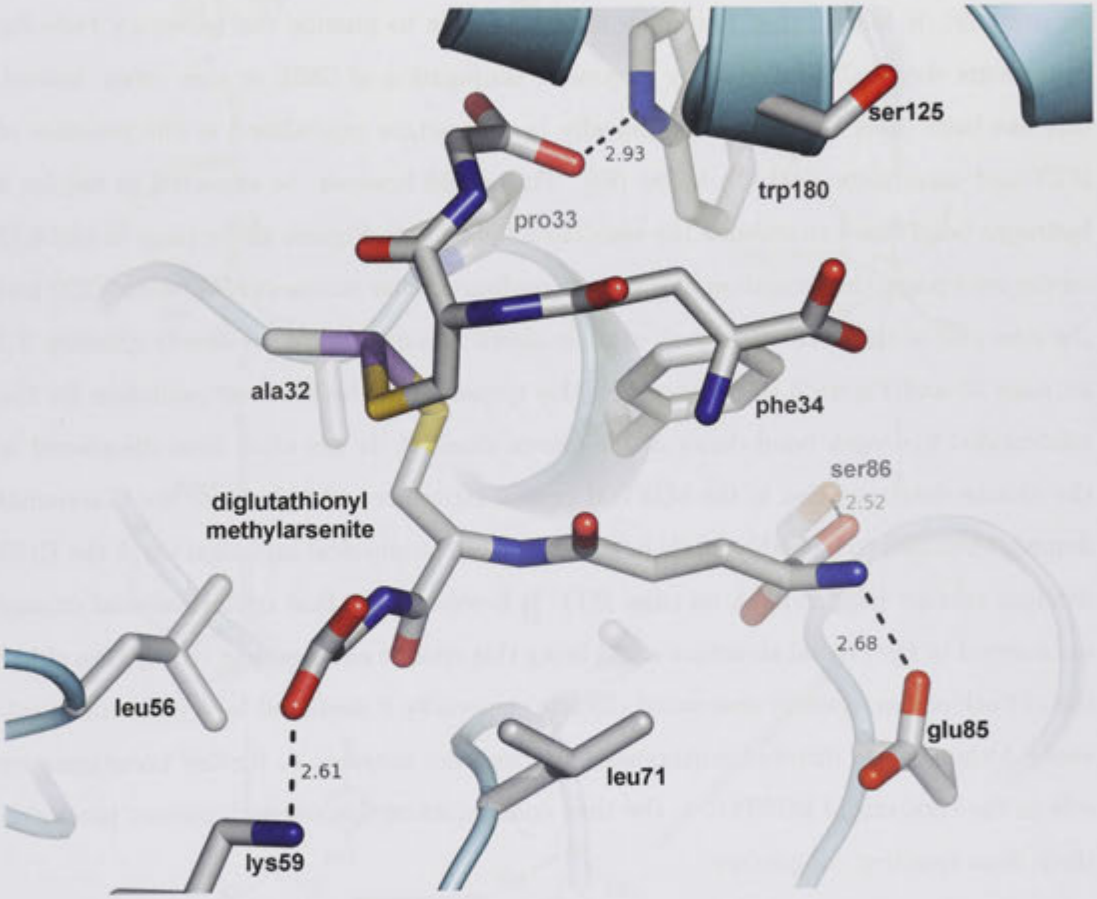


Figure 4.11

Diglutathionyl methylarsenite, refined to occupy approximately one third of the crystal lattice, is shown in stick representation along with its associated side chain residues. Atoms are coloured according to atom type. Alternate confirmations of native glutathione and surrounding solvent molecules have been omitted for clarity. Distances of the observed polar interactions with the polypeptide are given in Å. The only novel instance of such is the acceptance of a hydrogen bond from the the indole amine of W180 by aγ-glutamyl carboxylate of the modeled ligand.

ability to conjugate glutathione to CDNB [17], and the isoform could be expected to retain this functionality in the context of methylated arsenicals. As Figures 4.12 on the facing page and 4.13 on page 84 propose, the active site cysteine is not necessary for ligation of glutathione to arsenic, only for its reduction from the +5 to +3 oxidation state via donation of its lone electron pair. Given that one or both of the powerful reducing agents; DTT and TCEP, were present in the crystallisation conditions (Table 4.2 on page 70 and Table 4.1 on page 69), it follows that they may have been able to provide the necessary reducing equivalents chemically, followed by enzymatic conjugation of GSH, or *vica versa*. Indeed, this has been observed crystallographically in a structure crystallised in the presence of DTT and cacodylate (DMA^V) buffer [88]. This would however, be expected to require a hydrogen bond donor to stabilise the associated thiolate. In Figures 4.6 on page 75 and 4.12 on the next page, the formation of a solvent mediated polar interaction between Y229 and the lone pair of the trivalent arsenic atom is shown. As mentioned previously (Section 3.3 on page 50 and Figure 3.10 on page 59), this tyrosine may be the best candidate for the unidentified hydrogen bond donor of the omega class [4]. It has often been disordered in the routine determination of the hGSTO1 crystal structure and belongs to the C-terminal domain identified to be highly labile in structural and biophysical investigation of the E155 deletion mutant (Appendix A on page 181). It is conceivable that conformational change unobserved in the crystal structure could bring this residue close enough to stabilise either the glutathione or cysteine associated thiolate, especially if mediated by a molecule of solvent. Although site directed mutagenesis studies were initiated to further investigate its role in the catalysis of hGSTO1-1, the time constraints of this research project precluded them from reaching completion.

As mentioned above, the mutation of the active site thiolate to an alanine methyl group actually enhances the ability of hGSTO1-1 to conjugate glutathione to CDNB, why then, is a reaction intermediate observed in Figure 4.8 on page 77, and not the product shown in Figure 1.8 on page 12, 1-glutathionyl-2,4-dinitrobenzene? Given that within the structure presented in Section 4.3.1 on page 71, glutathione was not present in the initial co-crystallisation conditions (Table 4.1 on page 69), perhaps the most likely sequence of chemical events is those presented in Figure 4.14 on page 86. Here, the GSH and CDNB constituents of the soaking solution are hypothesised to have reacted in an enzymatic, or non-enzymatic, fashion to yield 1-glutathionyl-2,4-dinitrobenzene. Monomethylarsenite, previously reduced from MMA^V via chemical reduction by DTT, may have then displaced

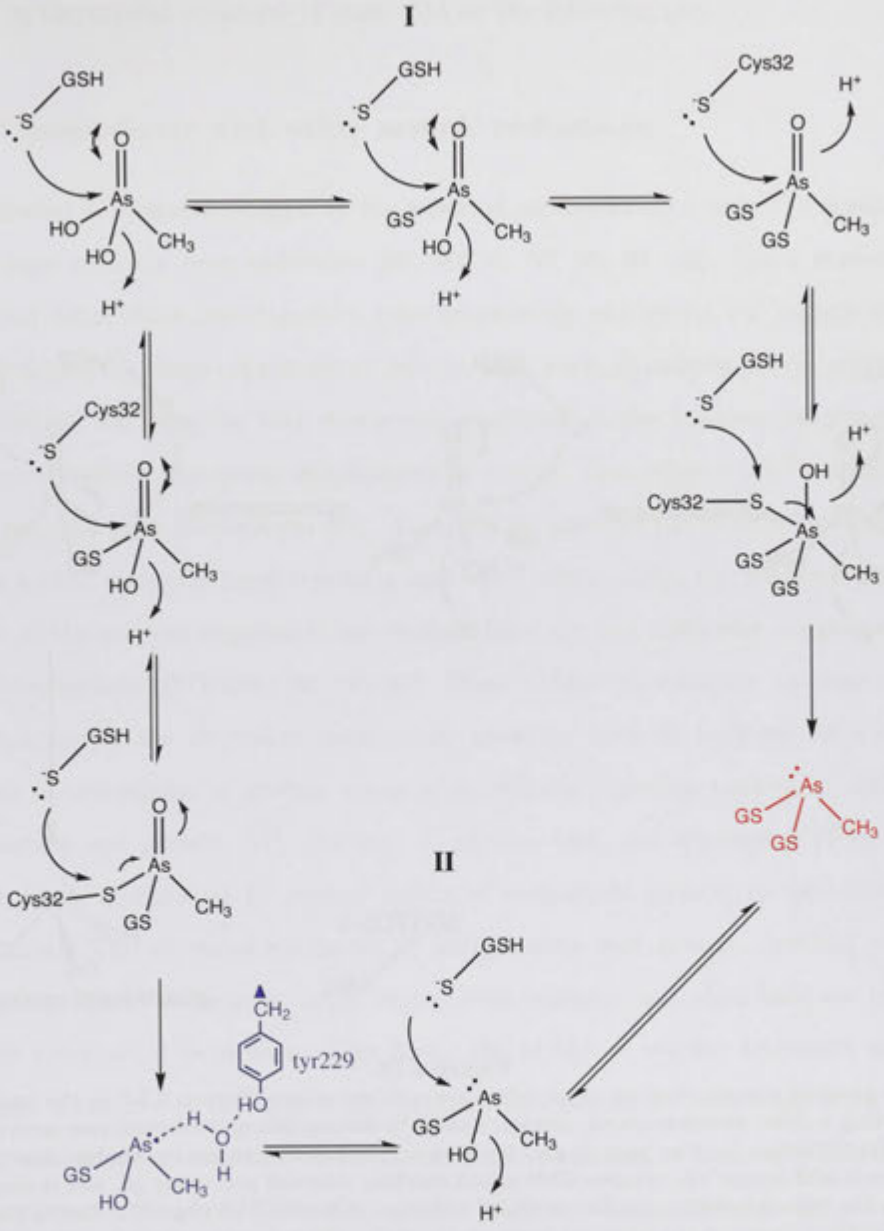


Figure 4.12

The proposed mechanism for the enzymatic formation of glutathionyl methylarsenite (blue) and diglutathionyl methylarsenite (red) by hGSTO1 native. These are the species modeled crystallographically, shown in Figure 4.5 on page 74 and Figure 4.10 on page 80 respectively. Two alternate pathways are given for the formation of the latter, but given the steric bulk associated with the penta-coordinated intermediate in scheme I, the latter is expected to be more physiologically accurate. Although within scheme II, the sequence of reactions is shown as glutathione conjugation, reduction and another conjugation, a sequence of two successive glutathionyl conjugations followed by a final reductive step with the active site cysteine could also occur. This mechanism is consistent with previous characterisation of microbial arsenate reductase enzymes (Section 4.3.4 on page 85), that show a direct interaction between the enzymatic thiol and the arsenic center. However, this reaction could likewise proceed by a deglutathionylation mechanism (Figure 4.13 on the next page) that would be more consistent with its mechanism of 4NPG reduction (Chapter 5 on page 97).

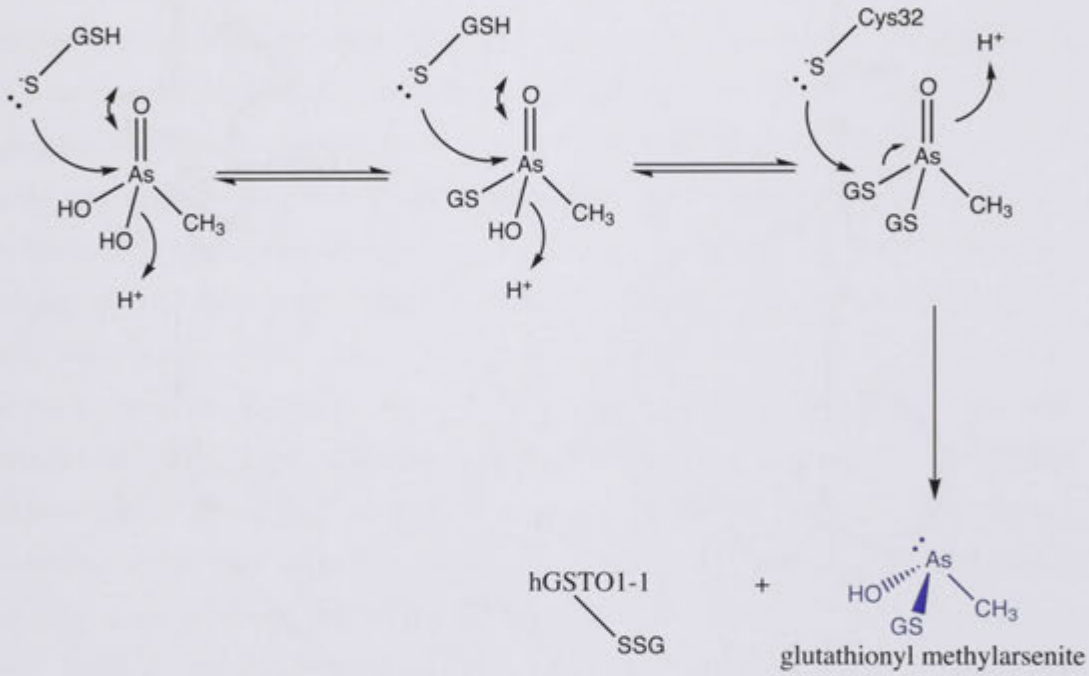


Figure 4.13

Despite previous characterisations of microbial arsenate reductases (Section 4.3.4 on the next page) observing a direct cysteine-arsenic covalent bond, the distinguishing thioltransferase activity of hGSTO1-1 (Section 1.1.3 on page 8) may be more accommodating to the mechanism shown here. This would still feature the cysteine-GSH mixed disulfide observed previously [4], and is analogous to the hypothesised mechanism of 4NPG reduction (Chapter 5 on page 97). Due to the deglutathionylation of the substrate however, the diglutathionyl intermediate hypothesised by Hayakawa *et. al.*, (Figure 4.1b on page 63) [71] would be precluded.

the glutathionyl functionality with one of its hydroxy ligands, yielding the chemical species observed in the crystal structure (Figure 4.14 on the following page).

4.3.4 Comparisons with other arsenic reductases

Several studies of enzymes integral to the assorted mechanisms of microbial arsenic detoxification have recently been published [89, 90, 91, 92, 93, 94, 95]. Using structural and biochemical data, these investigations have successfully elucidated the variety of mechanisms developed by these organisms to survive high loads of environmental arsenic, often in great detail. Perhaps the best characterised enzyme is the arsenate reductase (ArsC) of the gram-positive bacterium *Staphylococcus aureus*, that reduces iAs^{V} to iAs^{III} via a coupled reaction with thioredoxin [92]. Encoded on plasmid pI258 with two other genes encoding a iAs^{V} transcriptional regulator and iAs^{III} efflux pump, the sequence and crystal structure of the enzyme suggests it has evolved from the low molecular weight protein tyrosine phosphatases (PTPases) [96, 93, 92]. These ubiquitous enzymes are responsible for the dephosphorylation of post-translationally modified tyrosine residues via a cysteinyl-phosphate intermediate, a process essential in cellular signaling pathways, metabolism, differentiation and growth [97]. Indeed, *S. aureus* ArsC still displays PTPase activity, albeit at levels diminished by several orders of magnitude relative to its contemporary counterparts. The chemical similarity of phosphorous and arsenic, residing adjacently within group 15 of the periodic table, means that arsenate and phosphate are highly isomorphous tetrahedral oxyanions. This limits the ability of cellular transport systems to distinguish between the two and is the primary mechanism by which iAs enters the cell [89]. An evolutionary adaption to its detoxification appears to have assimilated the dephosphorylation scaffold of PTPases as a redox enzyme. Once reduced to the +3 oxidation state, the toxicity of the species increases markedly, but allows its active extrusion from the cell [92, 96]. Both PTPases and ArsC share the characteristic C-X_n-R active site motif of hGSTO1 (Figure 3.8 on page 56), but this appears to have arisen via a mechanism of convergent evolution to this class of reaction, since in all other respects, their structure and mechanism are distinctly different. Using several active site mutants and specific redox conditions during purification and crystallisation, the structure of *S. aureus* ArsC has been solved in all states of its reaction cycle, and unequivocally confirmed to function via an intramolecular disulfide cascade [93]. After initial attack by C10 upon the arsenic centre, expulsion of a hydroxy ligand as solvent is followed by subsequent formation of a disulfide

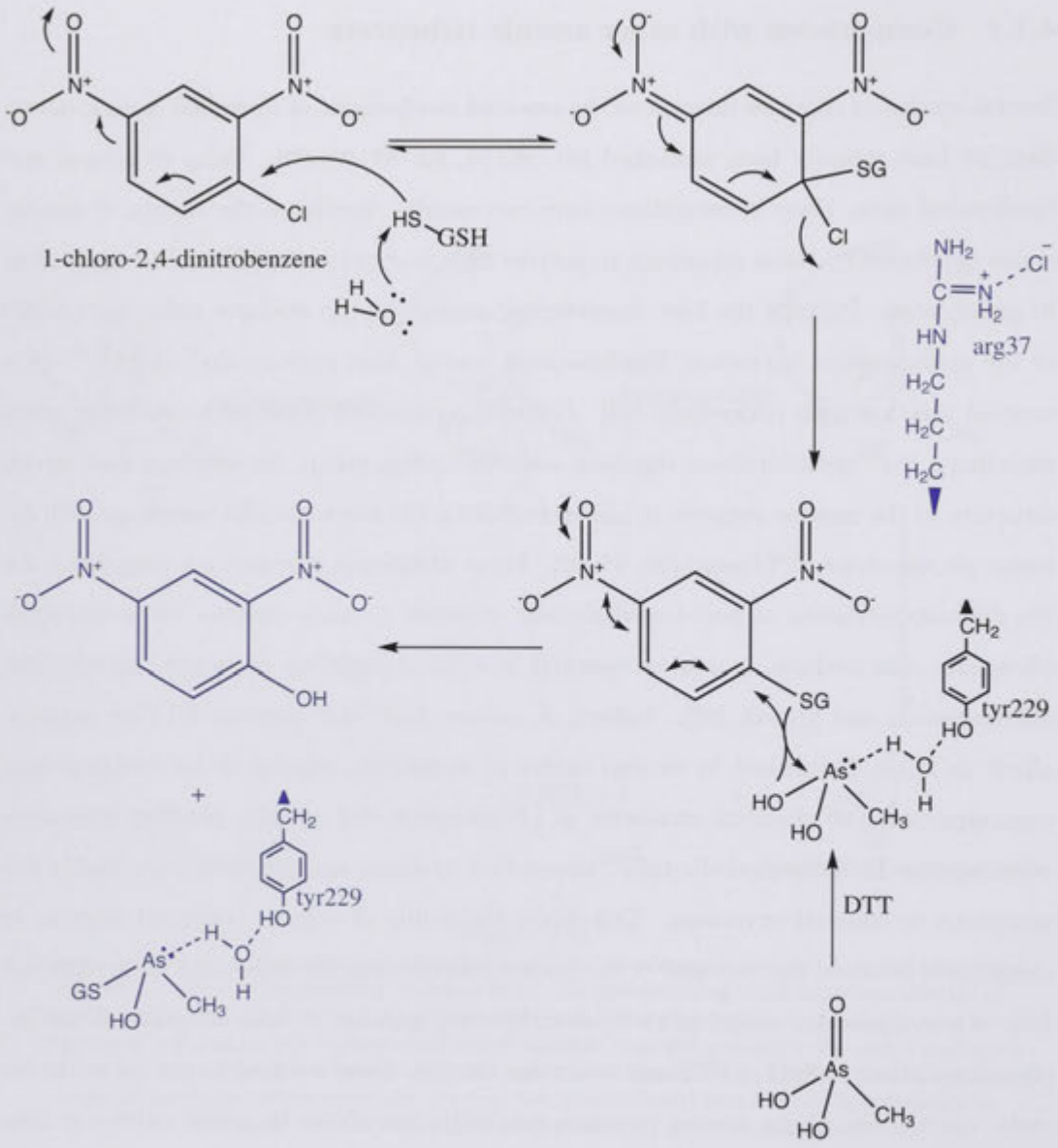


Figure 4.14

The sequence of chemistry hypothesised to have produced the chemical species presented in Figure 4.4 on page 73. The crystallographically observed intermediates are highlighted in blue. HDNB was likely observed in this experiment due to the enzymatic, or non-enzymatic, formation of 1-glutathionyl-2,4-dinitrobenzene, followed by displacement of its glutathionyl functionality by methyl arsenite, chemically reduced to the +3 oxidation state by the DTT present in the experimental conditions.

with nearby C82. This shuttles the electrons from the As-S bond to arsenic, concurrently releasing arsenite in the process. A major conformational change is then associated with C89 moving over 10 Å to form a C89-C82 disulfide, resulting in the regeneration of the reduced C10 thiolate, and exposure of the new disulfide at the solvent interface. This effectively shuttles the disulfide to a position accessible to reduction by thioredoxin, which can in turn be reduced by thioredoxin reductase and its co-substrate, NADPH (Nicotinamide adenine dinucleotide phosphate) (Figure 4.15 on the following page) [96, 93]. Of particular interest, a covalent C10-arsenic adduct within a C82S mutant could not be observed crystallographically due to a futile reaction cycle with solvent (Step 1 in Figure 4.15 on the next page). The reversibility of this initial step in the reaction cycle was confirmed unequivocally by observation of solvent derived ^{18}O incorporation via mass spectrometry [93]. Such a mechanism may have similarly hindered attempts to observe the equivalent methyl arsenic intermediate in complex with native hGSTO1 via co-crystallisation and soaking experiments involving MMA^V and DMA^V during this research. The pI258 ArsC C10-Arsenic intermediate was finally able to be observed by using a C15A (non-catalytic residue) mutant that was crystallised with 10 mM arsenite (+3 oxidation state) in the absence of DTT. The resulting crystal structure showed trivalent arsenic bound in a planar configuration perpendicular to the S-As bond [93]. A molecule of water was observed on the other side of the $HAsO_3^-$ plane representing a hydroxy leaving group, and forming the other apex of a trigonal bipyramidal prism. The other two catalytic cysteines were observed engaging in a disulfide bond [93]. This co-crystallisation with high concentrations of trivalent product most likely resulted in the observation of the cysteinyl-arsenic adduct crystallographically, due to an artificial shifting of the reaction equilibrium within the crystal to favour transition state complexation. At the time the equivalent experiments with hGSTO1 were being performed, trivalent methyl arsenicals were unavailable commercially, making the application of the same strategy difficult.

The two other families of arsenate reductases are represented by the ArsC enzyme of gram-negative *E. coli* encoded on plasmid R773 [91, 95], and Acr2p from the eukaryote *Saccharomyces cerevisiae* (yeast) [90]. Although sequence similarity suggests they are not related by evolution, both feature a mixed disulfide mechanism involving an active site cysteine residue and glutathione, similar to hGSTO1. The unicellular arsenate reductases differ from the omega class in that they require a ternary complex with glutaredoxin (Grx) to regenerate the active enzyme. Of the two, the *E. coli* enzyme has been most thoroughly

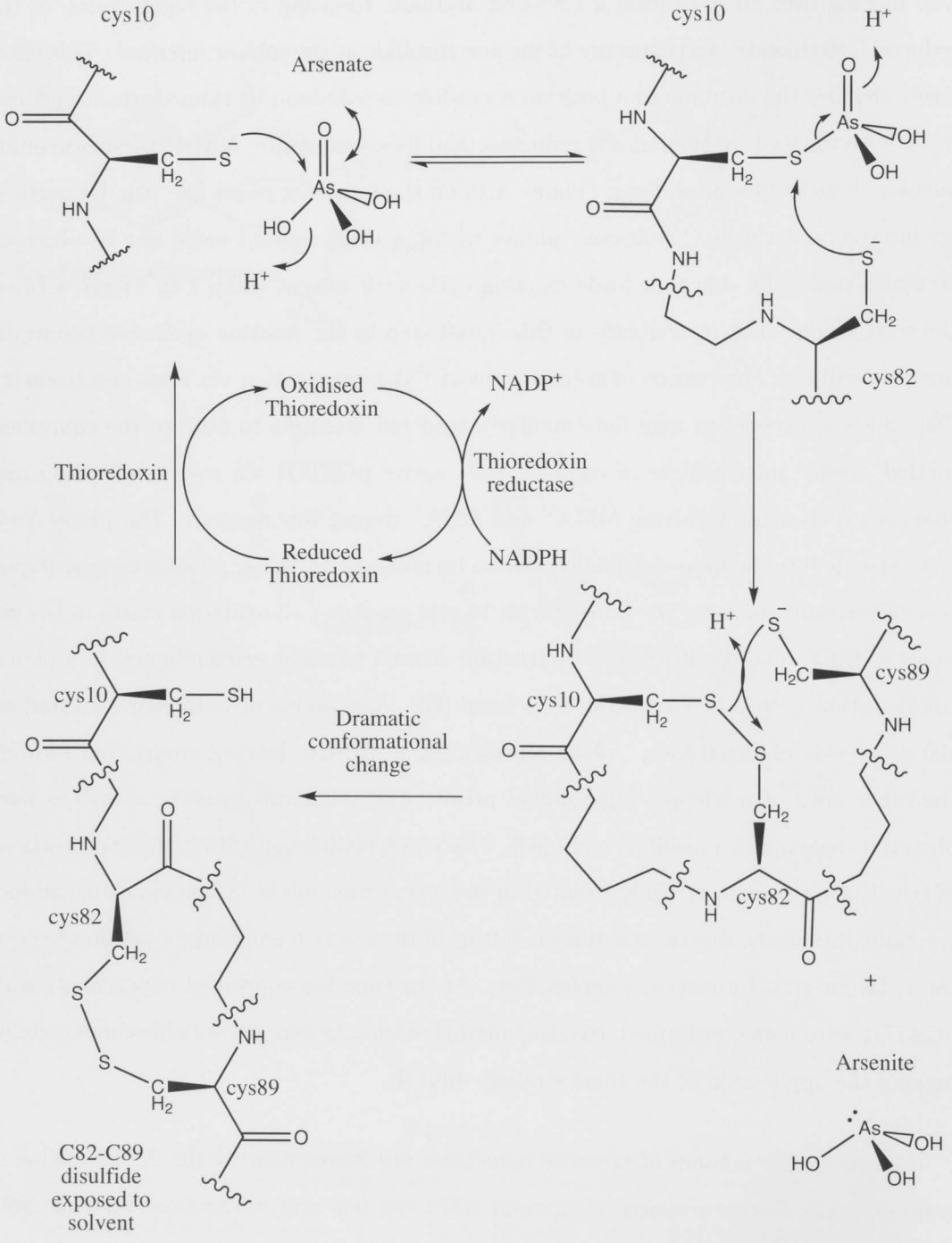


Figure 4.15
The mechanism of *S. aureus* arsenate reduction as elucidated by Messens *et. al.*, [93].

characterised, and the cysteine at position 12 found to be essential for catalysis [91]. Although any of the 3 glutaredoxin enzymes of *E. coli* are capable of participating in the reaction, Grx2 has the highest affinity for the ArsC-Arsenate-GSH complex [95]. All three of these substrates are required for interaction with Grx to occur, however substitution with substrate analogs such as sulfate or S-hexyl glutathione also allow the ternary complex to form, albeit with decreased affinity. Similarly, a C12S mutant is also capable of complex formation, indicating that free thiols associated with ArsC and GSH are desirable, but not essential, for formation of the ternary complex to occur [91]. Only GSH was capable of catalytic turnover however, and was unable to be substituted by the free thiol functionalities of β-mercaptoethanol or DTT [91]. The structures of *E. coli* ArsC and Grx2 have been solved by X-ray crystallography [95] and NMR spectroscopy [98] respectively, and their alignment with the scaffold of the hGSTO1 glutathionyl methylarsenite complex (Figure 4.4 on page 73) via the secondary structure matching (SSM) algorithm of Coot [43] reveals some striking similarities (Figure 4.17 on page 92). While the glutaredoxin fold is canonical of the GST family, the structural commonality of ArsC to the GST family was first noted during the preliminary NMR analysis of its structure [99]. It possess no detectable sequence similarity with hGSTO1, even within the segments of common secondary structure. The $\beta_1 - \alpha_1 - \beta_2 - \beta_3 - \beta_4 - \alpha_3$ motif aligns strongly (rmsd 2.32 Å) with the 140 residue fold of ArsC, however the latter contains a helix-turn-helix-loop-helix-loop-helix domain (residues 35 to 93) that replaces helix α_2 between β-strands 2 and 3. The R773 ArsC also features a short anti-parallel loop-strand-loop-strand-loop motif at its N-terminus that is absent in the omega class. Of most significance, this alignment results in an almost exact correlation between the active sites of the two enzymes (Figure 4.17 on page 92), with a distance between the superposed arsenic centers of just 4 Å.

In contrast, the 215 residue Grx2 displays 21% and and 45% sequence identity and similarity with hGSTO1 respectively (Figure 4.16 on page 91) [100]. This is reflected in the highly similar topological arrangement of Grx2 secondary structure, the only exceptions being the unique N- and C-terminal extensions of hGSTO1, and the Grx2 insertion of a helix-turn-helix motif (residues 100 to 139) between α-helices 4 and 5. The two structures superpose with a core rmsd value of 2.82 Å over 166 residues using the SSM algorithm of Coot [43].

The congruency of structural motifs between these enzymes could suggest a mechanism of hGSTO1 evolution that involved an ancestral gene fusion, effectively combining

the functionality of the *E. coli* ArsC-arsenate-Grx ternary complex into a single enzyme. The lack of a detectable sequence similarity between the R773 ArsC and hGSTO1 however, is suggestive that a stronger case may be argued for divergence from an ancestral glutaredoxin gene that arrived upon the structural features ideal for arsenate reduction by a process of convergent evolution instead. Indeed, there is no topological characteristic of hGSTO1 mirrored by ArsC that is not also present in Grx2, suggesting the glutaredoxin scaffold possesses all the features necessary to evolve into an arsenate reductase directly. For reasons that are unclear, gram-negative bacteria may have developed an evolutionary advantage in dividing their arsenate reductase mechanism between enzymes in the past. If so, this adaptation appears to have been lost in the subsequent evolution of mammals, who concurrently acquired a system of methylation. The enzyme identified to perform the equivalent reduction of iAs in humans, purine nucleoside phosphorylase (PNP), is a trimer that also shows no sequence similarity and limited structural congruency to the R773 ArsC. In any case, the complementary alignment of ArsC with hGSTO1 suggests that features of their inorganic and methylated arsenical reduction mechanisms may be conserved.

The proposed mechanism of R773 ArsC is shown in Figure 4.18 on page 94 [95, 101]. While roughly analogous to that shown for hGSTO1-1 in scheme II of Figure 4.12 on page 83, there are several distinctive differences. Primarily, the glutathione is shown to only participate in coordination to arsenate, and does not form a mixed disulfide residue as observed in hGSTO1 [4]. Secondly, although the arsenate-GSH-Grx binding sequence to ArsC has been well established by fluorescence experiments involving the artificial insertion of tryptophans within the ArSC active site [91], the precise order of initial nucleophilic attack by these thiols within the active site of hGSTO1 is still unknown. Given the close proximity of 2 tryptophan residues within the active site (W180 and W222), it is possible that future research focused on this area of hGSTO1-1 catalysis could elucidate this information by applying a similar strategy. Indeed, we have recently used the same methods to study the refolding kinetics of the hGSTO1Δ155 mutant (Appendix A on page 181). While initial attack of the MMA^V arsenic centre could either be by glutathione or cysteine, Figure 4.12 on page 83 hypothesises that the penultimate step of hGSTO1 catalysis involves attack of a glutathione on the active site cysteine to release a glutathionylated arsenical. This is contrasted to the mechanism presented here by DeMel *et. al.*, (Figure 4.18 on page 94), where Grx attacks the sulfur of a glutathione molecule, to release non-glutathionylated arsenite. The hGSTO1-1 equivalent of this mechanism is shown in

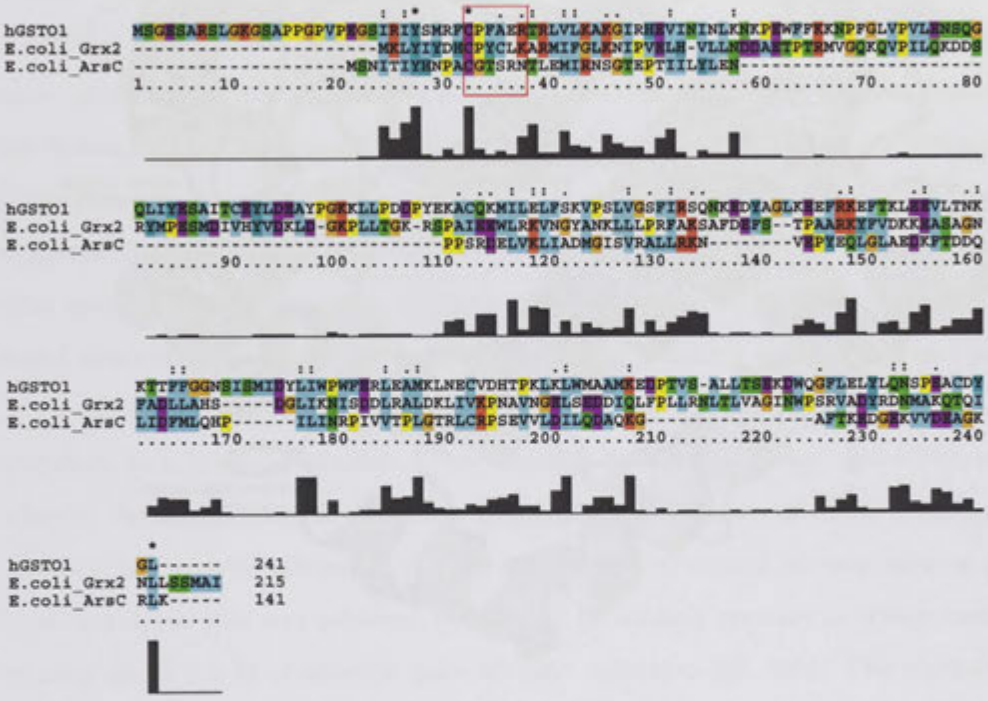


Figure 4.16

The multiple sequence alignment of hGSTO1 with the ArsC/Grx2 arsenate reductase couple of *E. coli*. The sequences corresponding to the hGSTO1 C-X_n-R active site motif is highlighted with a red box. While the blastp program [100] showed Grx2 to posses 21% and 45% sequence identity and similarity with hGSTO1 respectively, no similarity with the sequence of ArsC or Acr2p was detectable. In contrast, ArsC and Acr2p did show 23% and 42% sequence identity and similarity with each other. This would suggest that the striking correlation of ArsC's secondary structural motifs (Figure 4.17 on the next page) with parts of the hGSTO1 crystal structure has arisen from convergent evolution with efficiency to the arsenate reductase reaction acting as a selection pressure. This figure was generated using the program ClustalX [60].

Figure 4.17

The SSM superpositions of hGSTO1 (cyan) with the ArsC/Grx2 (magenta and yellow respectively) enzyme couple responsible for the reduction of arsenate to arsenite within *E. coli* is shown in the cartoon representation. The structure of the hGSTO1/methyl glutathionylarsenite complex (Section 4.3 on page 71) was first aligned with that of R773 ArsC (PDB identifier: 1J9B) [95], while the solution structure of Grx2 (PDB identifier: 1G7O, ensemble conformer 21) [98], was intern aligned to hGSTO1. The thiarsahydroxy derivative of the ArsC active site cysteine and the two molecules of arsenite observed in the crystal structure are shown in stick representation, coloured according to atom type. Overlaid in translucent stick representation is the position of methyl glutathionylarsenite shown in Figure 4.5 on page 74, demonstrating a conserved active site location. Elements of hGSTO1 secondary structure are labelled for reference. The structural correlation between the GST and glutaredoxin enzymatic families is well known, and superposed with a rmsd value 2.96 Å. Still, the 215 residue Grx2 enzyme features a helix-turn-helix motif (residues 100-139) inserted at the position corresponding to the hGSTO1 α-helix 4-5 junction. Of most interest however, is the high coincidence of secondary structure between ArsC and the N-terminal domain of hGSTO1, despite the two enzymes sharing no detectable sequence similarity (Figure 4.16 on the previous page). The $\beta_1 - \alpha_1 - \beta_2 - \beta_3 - \beta_4 - \alpha_3$ motif of hGSTO1 aligns strongly (rmsd 2.32 Å) with the 140 residue fold of ArsC, however the latter contains an all α-helical domain (residues 35 to 93) that links β-strands 2 and 3. This could be evidence of a distant ArsC-Grx2 gene fusion, combining their functionality into a single enzyme from which hGSTO1 may have descended. It is more likely however, that convergent evolution acted upon an ancestral glutaredoxin gene to arrive at the fold most suited to the reduction of pentavalent arsenicals.

Figure 4.13 on page 84. The former is hypothesised to be a more likely feature of methyl arsenic reduction in mammals, since glutathionylated, trivalent methyl arsenicals have been shown to be the required substrate for hAS3MT [72, 71]. Another striking difference between the two mechanisms is the presence of the ArsC-thiarsahydroxy species, likely to be stabilised: $S_{cys12} - As^{+} - OH \rightleftharpoons S^{+}_{cys12} = As - OH$. This is corroborated by the correlation of the observed bond planarity and lengths with those of small molecules in the Cambridge structural database, both of which were known with certainty due to the high resolution of the ArsC crystal structure. Such an intermediate is favoured in ArsC by the close proximity of 3 arginines and a histidine residue that encourage the dissociation of its hydroxy ligand. The resulting positive charge is then stabilised by an adjacent sulfate [95]. The lack of many basic, proximal residues within the active site of hGSTO1 however, suggests that a similar intermediate would not feature as part of its catalytic mechanism. The species is hypothesised to expedite catalysis due to its inherent instability causing rapid dissociation from the active site, concertedly reducing covalent product inhibition. It's authenticity was recently confirmed unequivocally by mutating one of the proximal arginines to a lysine or alanine. The resulting decrease in charge and steric repulsions allowed the charge neutral dihydroxy cysteinyl arsenite to be observed with a predicted decrease in catalytic efficiency [95]. Much like the structural investigation of *S. aureus* ArsC catalysis, this was achieved both times by soaking crystals in cryoprotectant containing up to 0.4 M of arsenite prior to data collection [95, 101]. The crystallographic observation of the cysteinyl-glutathione adduct resulting from step 2 in Figure 4.18 on the following page has thus far eluded the authors, however *in silico* molecular modeling [101] suggests that glutathione would coordinate arsenic in a position roughly analogous to its mode of binding observed in superposed hGSTO1 (Figure 4.4 on page 73), but translated by several angstroms and rotated by 180°.

The Acr2p arsenate reductase of *Saccharomyces cerevisiae* has been less well characterised, although its ability to restore a positive phenotype to an ArsC knockout strain of *E. coli* is indicative that it has a high level of functional and structural homology [90]. Indeed, R773 ArsC and Acr2p share 23% and 42% sequence identity and similarity respectively [100]. Notably, cloned Acr2p did not require co-expression with any host proteins, the eukaryotic enzyme was able to participate directly with *E. coli* Grx enzymes during the reaction [90].

Of particular interest is that despite the lack of sequence similarity, like the ArsC

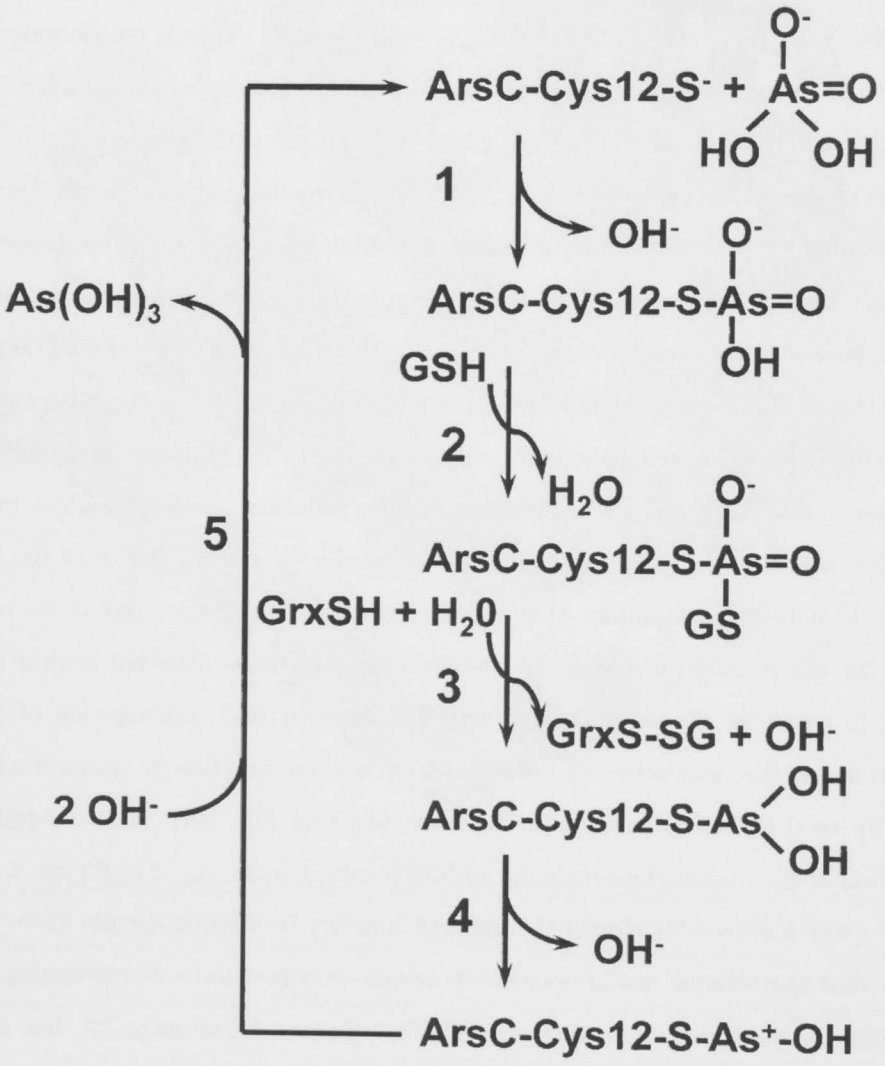


Figure 4.18

The mechanism of *E. coli* ArsC proposed by DeMel *et. al.*, [101]. While an analogous mechanism for hGSTO1 is shown in Figure 4.13 on page 84, the proposed mechanism of methyalted arsenic reduction shown in Figure 4.12 on page 83 is hypothesised to be more likely, since it is compatible with the release of glutathionylated products that are required by hAS3MT. This suggestes the would differ in the order of As-S bond formation and dissociation. In hGSTO1, it would be expected to be the Cys-As bond, rather than the GS-As bond shown here, that would provide the necessary reducing equivalents to the arsenic center.

enzyme of *S. aureus*, the *S. cerevisiae* Acr2p also appears to have descended from the PTPase family, but from the subgroup associated with the control of cell cycles [90, 93]. Identification of what appears to be an alternative branch of this evolutionarily pathway was recently made within the cyanobacterium *Synechocystis* sp. strain PCC 6803, that displays characteristics from both families of microbial ArsC enzymes [89]. This enzyme is a clear sequence homologue of PI258 ArsC and the alignment of 3 cysteines suggests it would also feature an intramolecular disulfide cascade. Unlike the resulting reduction by thioredoxin/NADPH in *S. aureus* however, the cyanobacterial enzyme was rather shown to be regenerated by an interaction with Grx and GSH, as in the mechanism of the R773 ArsC from *E. coli* and Acr2p [89, 95, 91].

The complex of hGSTO1 with S-(4-Nitrophenacyl)glutathione and oxidised glutathione:

Insights into the catalytic cycle afforded by the enzymatic complex of a specific substrate with bound product

The detoxification of α-haloketones to non-toxic acetophenones is an enzymatic reaction specific to hGSTO1-1 within the body [15]. After the α-halide is displaced by a chemical reaction with the high intracellular concentrations of GSH, recognition by hGSTO1 and nucleophilic attack of its active site cysteine residue subsequently liberates the corresponding acetophenone. Recently the novel compound S-(4-Nitrophenacyl)glutathione (4NPG) was synthesised and discovered to display a catalytic efficiency up to 15 times greater than conventional S-phenacyl glutathione [102]. It was a desirable outcome of this investigation to observe states within this catalytic cycle crystallographically to obtain a more fundamental understanding of hGSTO1-1 mediated chemistry. It was hoped that this would also provide insight into mechanistic commonalities with the other reactions it performs. Using crystals of the hGSTO1 C32A active site mutation, this research has succeeded in obtaining a high resolution structure of the ternary complex with this substrate and a reaction product, oxidised glutathione (GSSG). While the latter was observed within the active site, the binding mode of the substrate was unique, found to be residing at the dimer interface.

5.1 Introduction

α-Haloketones are biologically active compounds that can often enter the human body via various pathways. They have been identified as metabolites of certain insecticides, are used in laboratory research and as a temporary incapacitating agent in the form of 2-chloroacetophenone (tear gas). It is believed that a non-enzymatic attack by glutathione yields S-(phenacyl)glutathione derivatives that then act as substrates for enzymatic decomposition [15]. Now known to be mediated by hGSTO1-1, this mechanism is thought to proceed via nucleophilic attack of the active site cysteine upon the cysteinyl sulphur of the S-(phenacyl) glutathione, forming a mixed disulfide with the molecule of glutathione removed from the substrate. The resulting departure of a carboanion would be stabilised via enolization, yielding a corresponding acetophenone after protonation (Figure 5.1 on the facing page) [102, 15]. Physiologically, the native enzyme would be regenerated by nucleophilic attack of another glutathione molecule upon the resulting mixed disulfide, generating a reduced active site cysteine and oxidized glutathione, in a manner analogous to the reactions discussed previously (Section 3.3 on page 50 and Section 4.3 on page 71). This reaction does not have a specific requirement for glutathione, and its substitution for β-mercaptoethanol has been observed to increase the catalytic rate constant, k_{cat}, by a factor of five [15]. In contrast to all other activities of the Omega class, this reaction is unique to hGSTO1-1. hGSTO2-2 fails to show appreciable activity to this class of substrate [15]. A new compound, 4NPG, has recently been synthesised that has a turnover rate that is approximately 15 times higher, and displays a catalytic efficiency more than 200 times higher, than previously observed with S-(phenacyl)glutathione [102]. In addition, it allows hGSTO1-1 activity to be measured directly by a characteristic absorbance change at 305 nm. Previously, the enzymatic activity of hGSTO1-1 has been measured via an indirect spectrophotometric assay coupled to GSSG reduction by glutathione reductase. The concurrent oxidation of its co-substrate, NADPH, is associated with an absorbance change at 340 nm [102].

The specificity of this reaction for hGSTO1-1 is exemplified by experiments with human breast cancer cell lines. The T-47D cell line is natively deficient in hGSTO1-1 and was correspondingly unable to catalyse the reduction of 4NPG to 4-nitroacetophenone. In contrast, cells of the same lineage that had been transfected with a plasmid expressing recombinant hGSTO1 produced a robust absorbance change upon addition of substrate. Other breast cancer cell lines could similarly catalyse the reaction, effectively demonstrat-

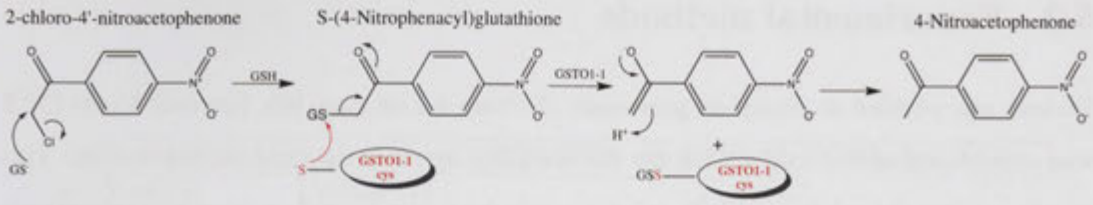


Figure 5.1
The proposed mechanism of 4NPG reduction by hGSTO1-1.

ing the dependence of the cell upon hGSTO1 for the mediation of this chemistry [102]. This presents for the first time, an assay that can identify the activity of hGSTO1-1 in human tissue as distinct from that of hGSTO2-2. The rate increase afforded by using β-mercaptoethanol as the thiol co-substrate is essential for successful characterisation of hGSTO1-1 activity within crude tissue extracts by this method [102].

The aim was to observe the binding mode of this substrate with hGSTO1 crystallographically, in the hope this would allow a structural insight into its specificity. The resulting structural information could also confirm the details of its reaction catenation and mechanism likely to be conserved within other aspects of its chemistry, such as the hypothesised thioltransferase modulation of the Ryanodine calcium release channel [28].

Soaking and co-crystallisation experiments with 4NPG involving crystals of hGSTO1 native consistently resulted in observation of the GSH-cysteine mixed disulfide as seen previously [4]. This was likely due to the formation of this reaction intermediate (step 2 in Figure 5.1) residing within a minimum of the reactions energy landscape. The subsequent dissociation of acetophenone is expectedly too rapid for observation crystallographically. Crystals of the hGSTO1 C32A mutant (Section 4.2 on page 66) were therefore employed in soaking experiments designed to discover the binding mode of 4NPG in its unreacted state within the active site. Yet the resulting crystal structures revealed an unexpected ternary complex, with 4NPG complexed along a crystallographic two-fold corresponding to the hGSTO1-1 dimer interface. In lieu of the substrate, the active site was complexed with oxidised glutathione (GSSG) with an occupancy=0.7, despite the absence of this compound in the experimental conditions (Table 5.1 on page 101). GSH was also observed at this site within the remainder of the crystal lattice. Although the same binding mode would likely also occur within crystals of the hGSTO1 C32S mutant, this was not confirmed experimentally in the current study. The insights into the enzymes catalytic cycle afforded by this high resolution crystal structures are discussed in more detail below.

5.2 Experimental methods

Protein was purified as described previously (Section 4.2 on page 66), however 5 mM DTT was substituted with 1 mM TCEP for the reducing agent in the final dialysis buffer. This was to prevent auto cleavage of the substrate in subsequent crystal soaking experiments via formation of a GSH-DTT mixed disulfide, in a manner analogous to the reaction with β-mercaptoethanol discussed above (Section 5.1 on page 98). Data was collected at the SSRL using a wavelength corresponding to 12 000 eV. In several equivalent experiments however, thermal motion in the glutathionyl portion of 4NPG prevented it from being modeled into electron density unambiguously. It was clear from inspecting the high resolution electron density of two such datasets that the mode of ligand binding in both datasets were identical, and that using information from both may facilitate a clearer image of the associated electron density. The two datasets were generated from crystals grown under similar conditions (Table 5.1 on the facing page) and collected within the same period of beamtime. They thus possessed coincident unit cell parameters, and could be combined using the 'Reflection file editor' utility within Phenix [45]. This effectively increased the observation to parameter ratio of the dataset, allowing the otherwise uninterpretable region of the electron density map to be modeled with substrate. While this did not alter any other structural details of the enzymatic fold, it did result in an artificially low number of modeled solvent molecules given the resolution (Table 5.1 on the next page). This is due to the position of many not being conserved, and their corresponding electron density peaks subsequently being averaged to zero.

The crystallisation and soaking conditions of each, along with data collection and refinement statistics are detailed in Table 5.1 on the facing page. Note that crystal G5 was previously co-crystallised with an inhibitor (Chapter 6 on page 131). Other crystals grown under identical conditions with this compound had consistently revealed an empty, unmodified active site, most likely due to the thiol reactive compound being unreactive with the hGSTO1 C32A active site mutant comprising these crystals. It was therefore deemed suitable for direct use in a 4NPG soaking experiment. Similarly, crystal G2 was first soaked in a GSH/reservoir solution before transfer to a drop containing 4NPG. This was envisioned to remove any residual TCEP from the storage buffer, to ensure that subsequently observed modifications of the substrate would be physiologically relevant. Despite these differences in methodology, the observed binding modes of 4NPG and GSSG in the resulting crystal structures were extremely similar, justifying the combination of these datasets.

Experimental parameters of the hGSTO1-C32A complex with 4NPG and GSSG		
Crystallisation	Crystal G2	Crystal G5
Reservoir solution	2.2 M $(NH4)_2SO_4$ 100 mM Sodium Acetate pH 4.25	2.2 M $(NH4)_2SO_4$ 100 mM Sodium Acetate pH 4.75
Crystallisation drop	1 µL hGSTO1 C32A (32 mg/ml)	1 µL hGSTO1 C32A (32 mg/ml) (containing 150 nm of the 'compound 8' inhibitor‡)
	1 µL reservoir	1 µL reservoir
Soaking solution a∓	2 µL reservoir 0.5 µL 100 mM GSH pH 7.5	2 µL 10 mM 4NPG pH 7.0 2 µL reservoir
Soaking solution b∓	2 µL 10 mM 4NPG pH 7.0 2 µL reservoir	
Cryoprotectant solution	2.75 M Li_2SO_4	2.75 M Li_2SO_4
	100 mM Sodium Acetate pH 4.75	100 mM Sodium Acetate pH 4.75
	Sequential transfer from 5%—->10%—->15% Glycerol solution	Sequential transfer from 5%—->10%—->15% Glycerol solution
Data collection		
Resolution limits, Å	50.0-1.81 (1.87-1.81†)	28.5-1.68 (1.74-1.68†)
Space group	$P3_121$	$P3_121$
Unit Cell, Å	a=b=57.63 c=140.1	a=b=56.95 c=139.8
Total reflections	237 715	210 572
Unique reflections	24 276	31 004
Completeness, %	96.0 (68.4†)	98.0 (84.2†)
I/σ	37.1 (3.2†)	34.9 (1.8†)
Redundancy	9.8 (5.9†)	6.8 (4.2†)
Refinement		
R_{cryst}/R_{free}, %	19.4/23.5	
rmsd bonds, Å	0.007	
rmsd angles, °	1.292	
Number of waters	91	
Ligands	$SO_4{}^{2-}$, glycerol, glutathione, GSSG, 4NPG	

†, the highest resolution shell. ∓, allowed to equilibrate for 48 hours. ‡, Please see Chapter 6 on page 131.

Table 5.1
Crystallisation details and structure determination statistics of the hGSTO1-C32A, GSSG, 4NPG complex.

5.2.1 The calculation of average 'kicked' maps

In order to reduce bias and noise, the phenix.maps utility [45] was used to generate average 'kicked' maps for visualisation of ligand binding. 'Kicked' maps involve introducing many small, random perturbations of the model coordinates, from each of which a map is calculated. The final map is an average of this series of perturbed maps, and effectively decouples any correlations in structure factors or chemistry terms (such as bond lengths and angles) that may be conveying phase bias (Section 2.5.5.2 on page 37). This technique was specifically employed in this instance both to ensure that either of the two combined datasets did not overly bias the result obtained, and because it has been shown to allow otherwise uninterpretable regions of electron density to be accurately assigned to the model [103]. This latter feature was particularly attractive given that regions of electron density corresponding to 4NPG and GSSG were difficult to decipher.

In more detail, the electron density maps are calculated:

$$AK = FT[\sum(2F_{obs} - k \times F_{model})] \tag{5.1}$$

in which each individual 'kicked' $|F_{model}|$ amplitude is scaled to its respective $|F_{obs}|$, before individual $2F_{obs} - F_{calc}$ maps are computed and added to the sum. The mean value of these functions then forms the final Average Kicked map (AK) [103]. Since the Fourier Transform (FT) is a linear operation, the process can be expedited computationally by replacing the real space averaging of maps with corresponding Fourier coefficients, followed by a final map calculation at the end.

5.3 Results and Discussion

The aim of the experiment detailed above was to visualise unreacted 4NPG in the G-site of hGSTO1. Instead, this position was found to be occupied by GSSG with an occupancy of 0.7, despite this compound being absent in the experimental conditions. Presumably, GSH oxidised to its homodimer spontaneously during the period of crystallisation. One half of the dimeric compound, extending upwards towards the C-terminal domain, interacts with residues of the H-site with its glycinyl group. The remaining 30% of sites within the crystal are occupied by reduced glutathione, in the same position observed in the native complex [4]. While the former is a product of all hGSTO1 mediated catalysis, this is the first time its binding mode to the enzyme has been confirmed crystallographically. An associated

conformational change has again occurred in the α6-α7 loop, but is of a different nature than observed previously (Figures 5.3 on page 106 and 5.5 on page 108).

The glycinyl moiety of 4NPG was observed just a few Å from the γ-glutamyl tail of oxidized glutathione, binding along the crystallographic two fold axis of symmetry corresponding to the dimer interface (Figure 5.2 on the following page). The nitrophenacyl functionality is observed pointing downwards into the dimeric cleft. This results in the cysteinyl sulfur of 4NPG residing 17 Å away from the c-β atom of A32, suggesting this is not a catalytically relevant binding mode, a notion that has significant historical precedent in the literature. In fact, the GST family of enzymes were originally known as 'ligandins', due to their apparent capacity to bind a wide variety of large (>400 Da) lipophilic compounds such as bile acids, fatty acids and certain drugs, in addition to their function as enzymes. It is thought this may play a role in storage and transport of these compounds in the aqueous phase of the cell [104]. The position of several of these 'ligandin' or 'L-site' binding pockets have been confirmed crystallographically in a broad spectrum of GSTs. While their positions within the ϕ class of *Arabidopsis thaliana* [105] and the human π class [106] were observed immediately adjacent to the active site, the binding of the anti-Schistosomiasis drug Praziquantel, to an α class GST of the parasitic worm *Schistosoma japonica*, was similarly found to occur at a position distal to the G-site, also within the dimer interface [107]. While this polyaromatic is dissimilar to 4NPG in structure, an analogous glutathione conjugate, S-(3-iodobenzyl)glutathione, was found to bind in a similar position within a σ class GST of squid [108] (Figure 5.9 on page 113). Despite these structures suggesting the associated binding modes were not catalytically relevant, the close proximity of 4NPG and GSSG in this instance, may suggest a concerted mechanism in which 4NPG may enter the active site as the latter departs. Indeed, the inability to observe both of these compounds in isolation suggests that their binding modes are correlated within the enzymatic cycle. This is supported by the binding of 4NPG being associated with rearrangement of the symmetrically related E91-K114 salt bridges, the only instances of such interactions between the subunits. The four residues now form a four way interaction at the apex of which is a solvent mediated contact to the phenacyl carbonyl of the substrate (Figure 5.8 on page 112). It is a possibility that such perturbation of inter-dimeric bonding may be able to infer binding cooperativity within the enzymes quaternary structure.

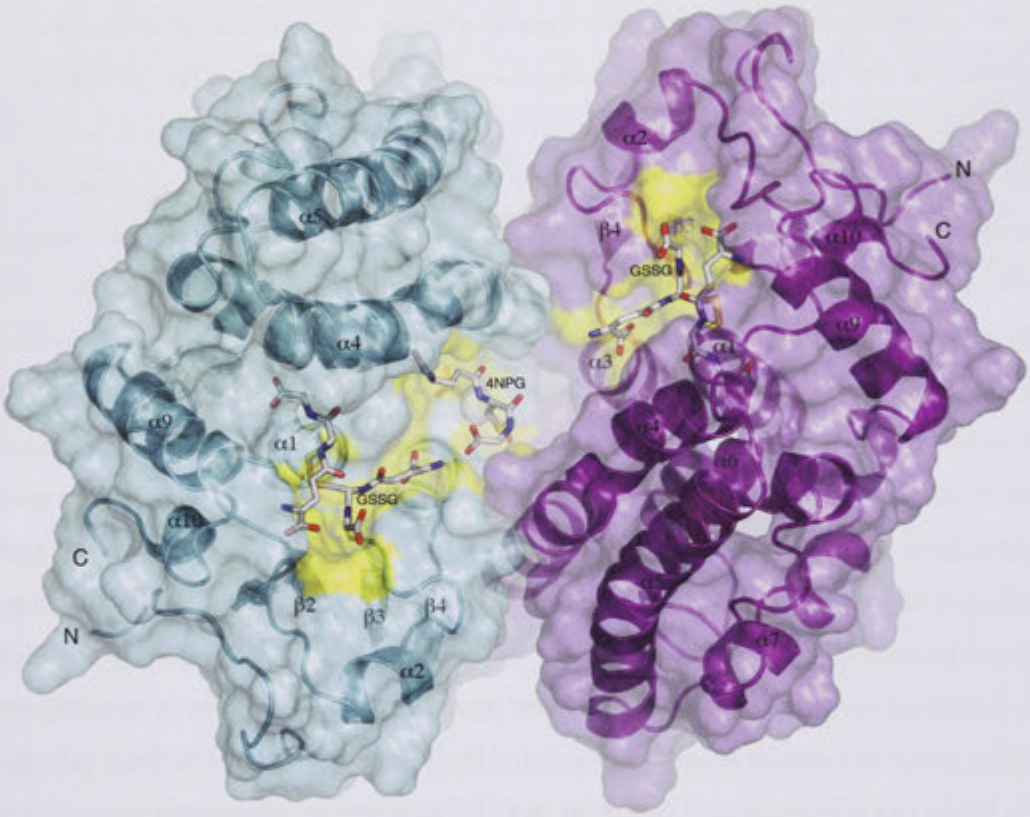


Figure 5.2
A view from above the hGSTO1-1 physiological dimer. Opposing monomers are shown in magenta and cyan cartoon representations covered by a translucent molecular surface. The ternary complexation with GSSG and 4NPG is shown in stick representation, with ligand binding highlighted in yellow. Whereas GSSG is observed to bind within the enzymatic active site, the interaction with 4NPG occurs along 2-fold axis of symmetry, that may correspond to a 'ligandin' binding mode observed in other GSTs [107, 104, 108, 105, 106]. The mutually inclusive binding of these ligands and their close proximity within the crystal structure however, suggests this may be a preliminary docking position of the derivatized xenobiotic, allowing it to move into the active site concertedly with GSSG departure.

5.3.1 Crystallographically observed mode of GSSG binding within the hGSTO1 active site.

The polar interactions with the half of GSSG occupying the G-site are analogous to those observed for reduced glutathione binding in the same position (Figure 5.4 on page 107). In contrast, interactions with the other half of the dimeric molecule, extending upwards into the H-site, are almost exclusively hydrophobic in character. Exceptions are a hydrogen bonding interaction between its γ-glutamyl carboxyl and the main chain carbonyl of L56, and an internal hydrogen bond between its γ-glutamyl carbonyl and the glycinyl amine occupying the G-site. This lack of defined interactions with the protein undoubtedly contributes to the relatively high *B*-factors and ill defined electron density associated with this H-site bound half of the molecule, suggesting significant thermal motion (Figure 5.3 on the following page). In addition, its complex formation is seen to be associated with the structural rearrangement of several amino acid side chains relative to the previously published complex with glutathione [4]. Of particular interest is the movement of Y229 to accommodate the γ-glutamyl residue, further highlighting the dynamic plasticity of this side chain within the hGSTO1 active site. The side chain of W222 has also adopted its alternate conformer, as seen upon binding of HDNB (Figure 4.8 on page 77), however the reoriented indole group is now too distant to engage in a hydrogen bonding interaction with the sulfur lone pair of M187. Although residues 183-196 have again responded to H-site complexation with conformational change, it is of a different nature than observed in the ternary complex of glutathionyl methylarsenite/HDNB (Figure 5.5 on page 108). In addition, the nearby side chains of K57, I131 and R132 have been disordered.

5.3.2 Crystallographically observed 4NPG binding at the hGSTO1-1 dimer interface

Rather than the expected binding within the G-site, 4NPG was observed to bind along the crystallographic 2-fold corresponding to the hGSTO1-1 dimer interface (Figures 5.2 on the preceding page and 5.6 on page 110). The molecule is observed orientated with its aromatic functionality pointing downwards into the dimeric cleft, with the glutathionyl portion of the molecule extending upwards towards the G-site. Its glycinyl carboxylate is seen to form a hydrogen bonding network to the γ-glutamyl amino acid functionality of GSSG via the side chain of E85 and two bridging molecules of solvent, while the γ-glutamyl carboxylate displaces two water molecules to form a salt bridge with the conserved guanidinium of R37

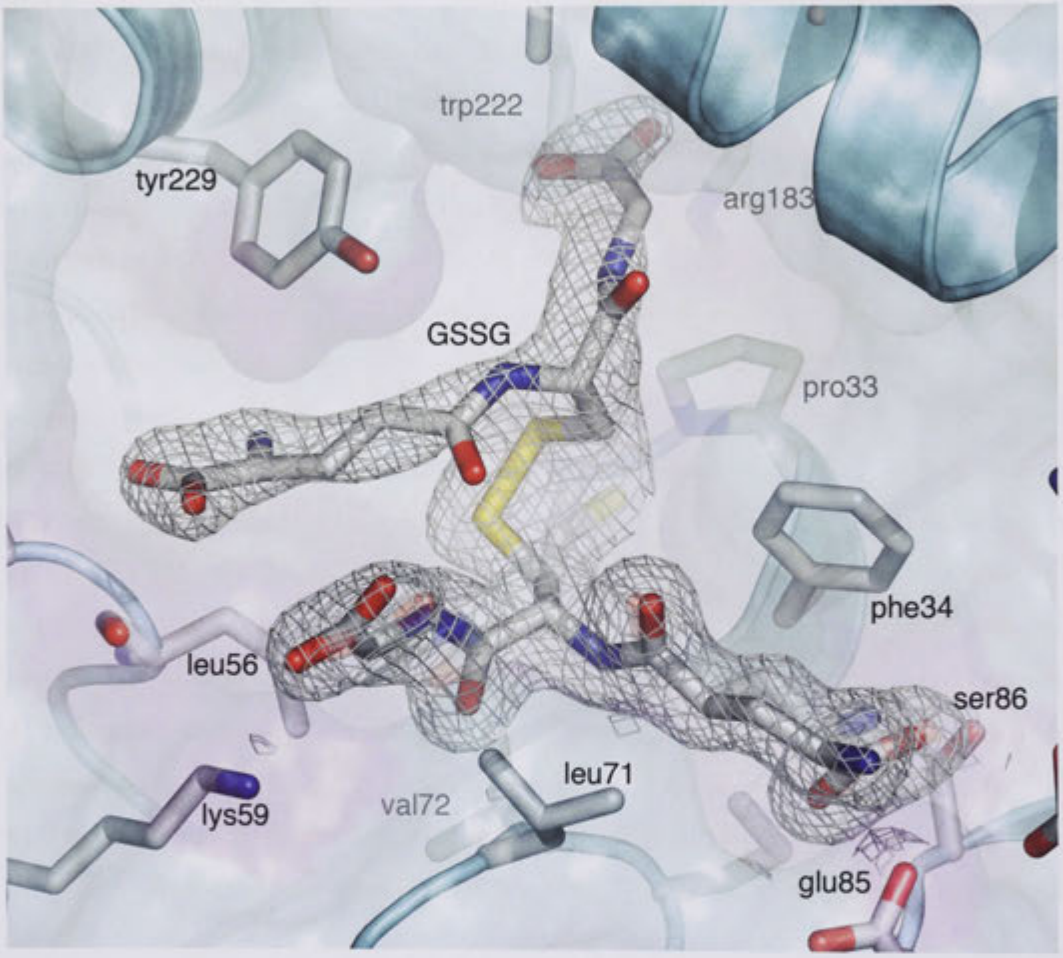


Figure 5.3
A closer inspection of the hGSTO1 active site showing the binding of GSSG and GSH in stick representation, the latter in translucent. These oxidised and reduced forms of glutathione were refined to occupancy values of 0.7 and 0.3 respectively. The weighted $2mF_O - DF_C$ electron density corresponding to GSSG is shown as gray isomesh contoured to a level of 1σ. To reduce distortion, bias and noise, the map shown has been 'kicked' using the phenix.maps program [45] (Section 5.2.1 on page 102). hGSTO1 is shown as a cyan cartoon overlaid in a translucent molecular surface. The interaction of GSSG with the polypeptide is highlighted in magenta.

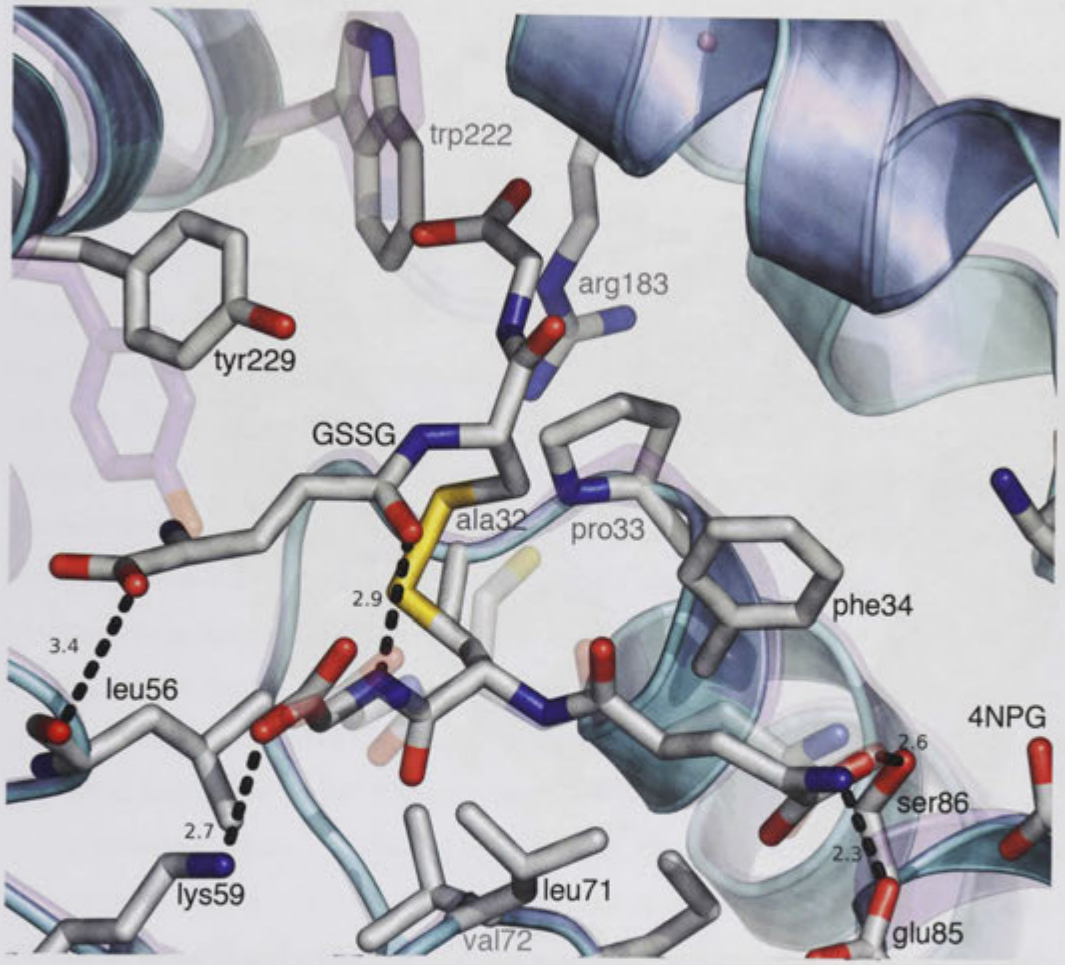


Figure 5.4

Oxidised glutathione and associated amino acids within the hGSTO1 active site are shown in stick representation, coloured according to atom type. GSH was refined to an occupancy of 0.3 and is overlaid in translucent stick representation within the G-site. While the corresponding half of GSSG shares the polar interactions of GSH previously observed [4], the other half, extending upwards into the H-site, exhibits interactions that are principally hydrophobic in character. Relative to the GSH bound structure [4], it's binding has been accommodated by a movement of the Y229 hydroxyl by 4.5 Å and the adoption of an alternate conformer of W222. The side chain conformations of nearby K57, I131 and R132 have been disordered (not shown). The crystallographically observed distance of all polar contacts are given in Å. The edge of nearby 4NPG is visible on the right hand side.

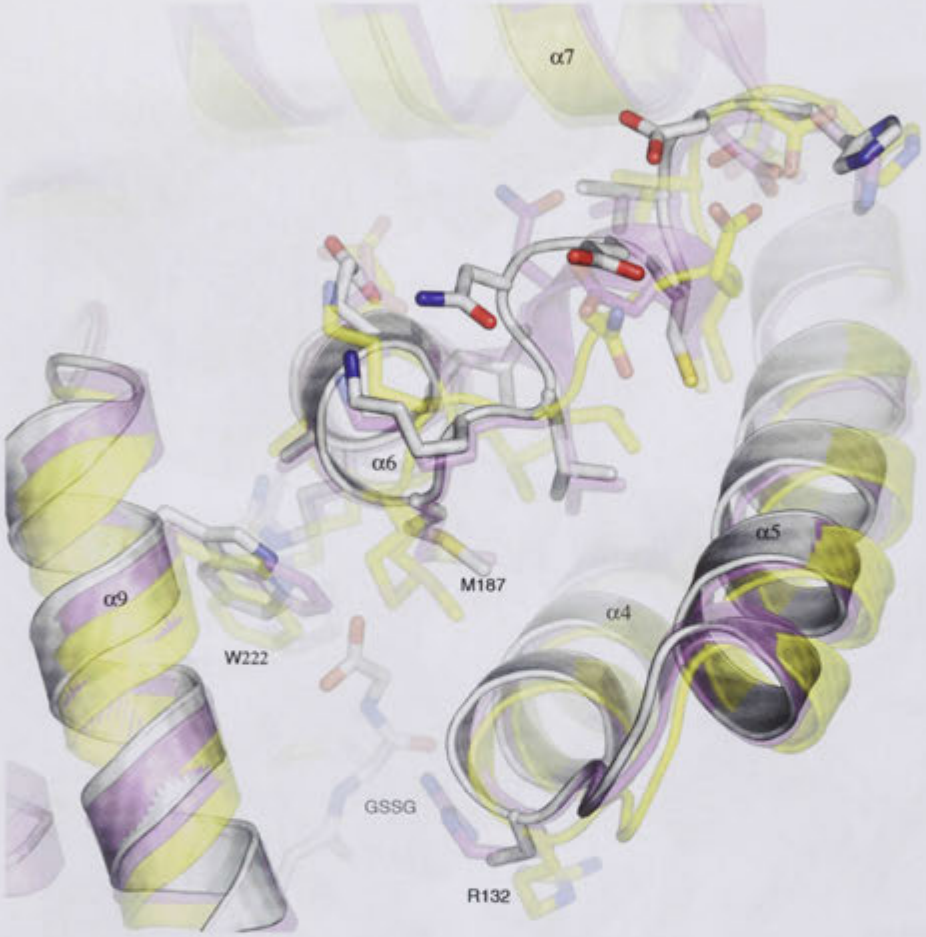


Figure 5.5

The conformational change within the α6-α7 loop associated with GSSG binding in the adjacent H-site (gray) is contrasted to that observed in the complex with glutathionyl methylarsenite/HDNB (yellow), and in the complex of the native enzyme with glutathione (magenta) [4]. One of the terminal glycine groups of GSSG is seen extending into the H-site in a position equivalent to the binding of HDNB observed previously (Figure 4.9 on page 78). The associated side chains of W222 and R132 have likewise adopted alternate conformations, however the latter is no longer stabilised by a sulfate mediated salt bridge and its guanadinium group is disordered. Furthermore, the alternate rotamer of W222 no longer forms a hydrogen bond with the sulfur lone pair of M187. This structural change is suggestive of a 'conformational switch' that is specific to ligand iterations within the H-site, since it does not occur in response to either GSH [4] or ascorbate (Section 3.3 on page 50) binding within the G-site. The structural plasticity associated with the transition of the α6 C-terminus from a kinked-helix, to a random coil loop region, is likely to be essential for the structural dynamics associated with ligand binding and efficient catalytic turnover. The reader is directed to Figure 4.9 on page 78 for cross reference.

(Figure 5.7 on page 111). Interestingly, its binding has also stimulated the rearrangement of the E91-K114 salt bridges, the only instances of polar interactions between the subunits. The side chains of both residues have moved several angstroms towards each other, forming a four way salt bridge (Figure 5.8 on page 112). At the apex of this interaction is a bridging molecule of solvent located precisely on the crystallographic two-fold, residing directly below 4NPG. This molecule of solvent concurrently accepts hydrogen bonds from the K114 amine and its symmetry related partner while donating one to the phenacyl carbonyl group of 4NPG. While the movement of E91 and K114 from the opposing dimer are enforced in this instance by the internal symmetry of the crystal, it is likely that they would follow their electrostatic counterparts upon binding of the substrate physiologically. Notably, this rearrangement of the only electrostatic interaction between monomers may represent a mode of communication and possible cooperation associated with this mode of catalysis. This is corroborated by both ligands being observed concurrently despite many attempts to visualize hGSTO1 complexed to each in isolation. During previous kinetic characterizations of this reaction [102] this possibility was not investigated, and future work may be warranted to allow an estimation of cooperativity.

The newly orientated carbon chain of K114 now forms the floor of a hydrophobic binding groove accommodating the nitrophenacyl functionality. The remaining walls of this cavity are comprised of the lipophilic side chains of A87, L103 and L176. M172 forms the end of this greasy pocket (Figure 5.7 on page 111).

As mentioned above (Section 5.3 on page 102), the binding mode of 4NPG presented here, may be representative of a 'ligandin' binding mode also observed in other classes of GST [107, 104, 108, 105, 106]. Two crystal structures exemplifying this warrant particular mention. The binding of the anti-Schistosomiasis drug Praziquantel to an α class GST of the parasitic worm *Schistosoma japonica* [107], and the complex formation of the glutathione conjugate, S-(3-iodobenzyl)glutathione with a σ class GST of squid [108] are both reminiscent of the dimer interface mode of binding that has currently been discovered for 4NPG. The crystal structures of these enzymatic complexes were superimposed with the 4NPG/GSSG ternary complex for comparison (Figure 5.9 on page 113).

5.3.3 Comparisons with other binding modes of GSSG

There is only two instances of GSTs complexed to GSSG within the PDB . The first is a representative of the human Mu class, hGSTM2-2 (PDB identifier: 1YKC) [109], while the

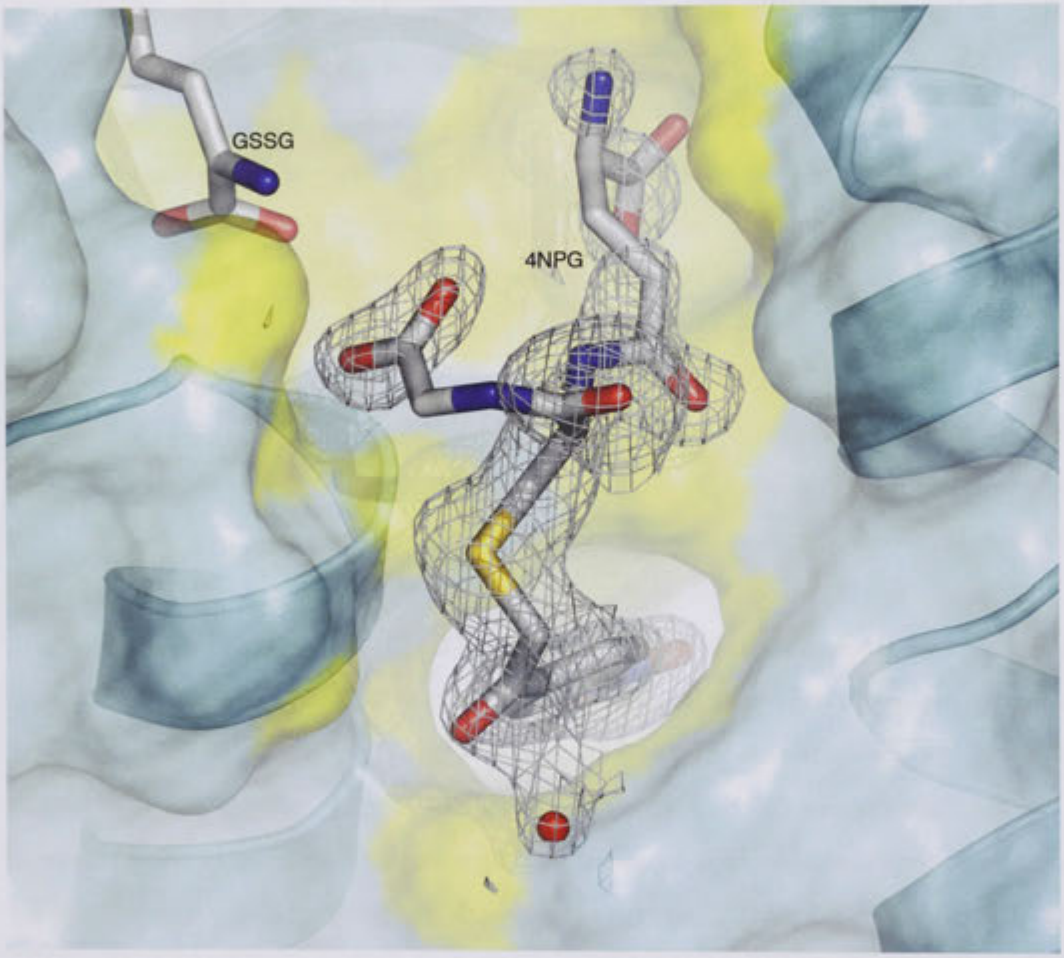


Figure 5.6

4NPG is shown in stick representation, coloured according to atom type. The corresponding region of electron density is shown as a weighted $2mF_O - DF_C$ map, that has been 'kicked' using phenix.maps utility [45] (Section 5.2.1 on page 102), depicted as gray isomesh contoured at a level of 1σ. A γ-glutamyl amino acid moiety of GSSG is visible in the upper left corner. The crystallographic fold of hGSTO1 is shown as cyan cartoon overlaid with a translucent molecular surface. As in Figure 5.2 on page 104, interactions with the bound compounds are highlighted in yellow.

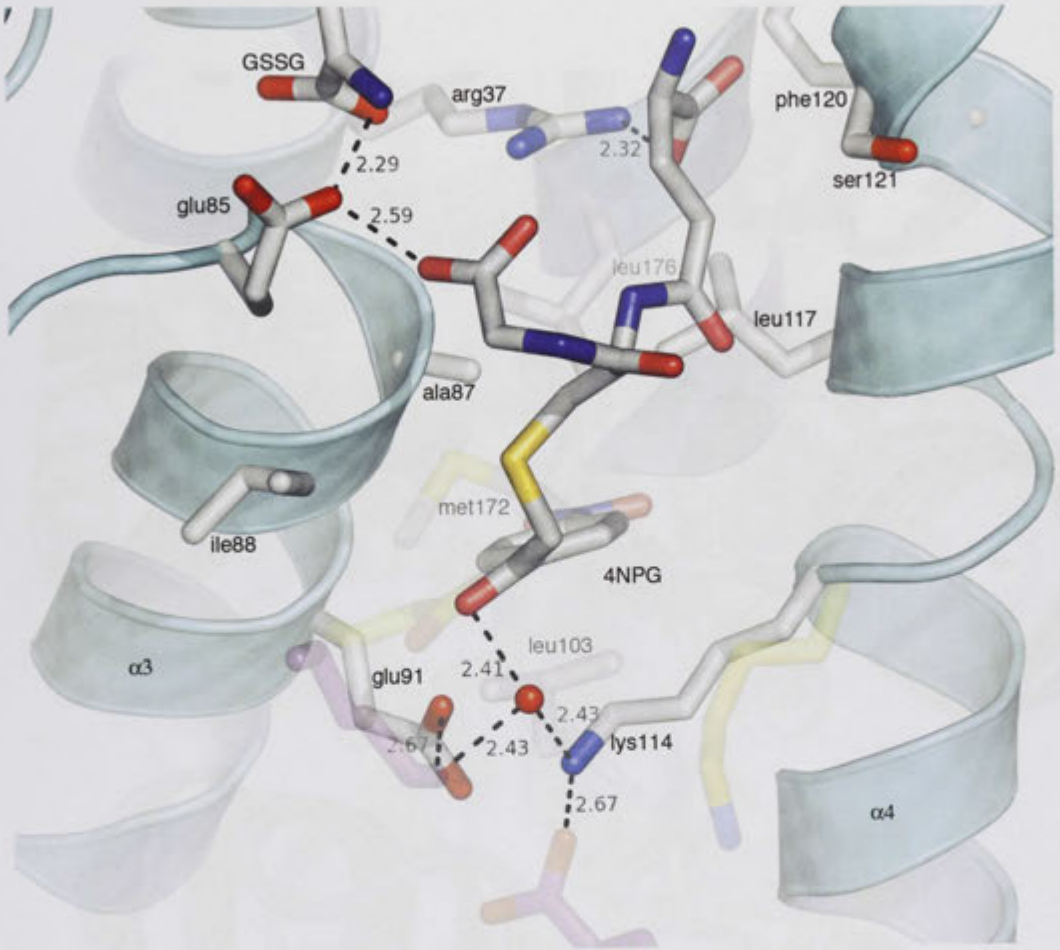


Figure 5.7

4NPG is shown in stick representation along with associated amino acid residues, coloured according to atom type. Of particular interest is the associated rearrangement of the E91-K114 salt bridges. The two residues have moved several Å towards each other to concurrently form a hydrogen bonding network to 4NPG via a bridging solvent molecule, that is located precisely on the crystallographic 2-fold. The symmetrically generated instances of these two residues are shown in translucent magenta, while their position in the glutathione bound native structure [4] are shown in yellow. The symmetrically generated copy of 4NPG is not shown for clarity. For a clearer view of this structural change, the reader is directed to Figure 5.8 on the next page. At the other end of the molecule, the glycinyl carboxylate is hydrogen bonded to a γ-gutamyl amine of GSSG via the side chain of E85 and solvent molecules (not shown). The γ-glutamyl end of 4NPG on the other hand, is observed to form a salt bridge interaction with the guanidinium of R37. The crystallographic fold of hGSTO1 is shown as a cyan cartoon.

Figure 5.8

A view of the E91-K114 salt bridge rearrangements approximately parallel to the crystallographic 2-fold axis of symmetry. The molecule of hGSTO1 in the asymmetric unit is shown as a cyan cartoon with its symmetrically generated equivalent in magenta. The new polar interactions surrounding 4NPG are shown with dashed lines, with crystallographically observed distances reported in Å. The position of the symmetrically related ion pairs observed in the native complex [4] are shown in stick representation overlaid in translucent yellow. For clarity, other side chains are shown only for the molecule within the assymetric unit, coloured according to atom type.

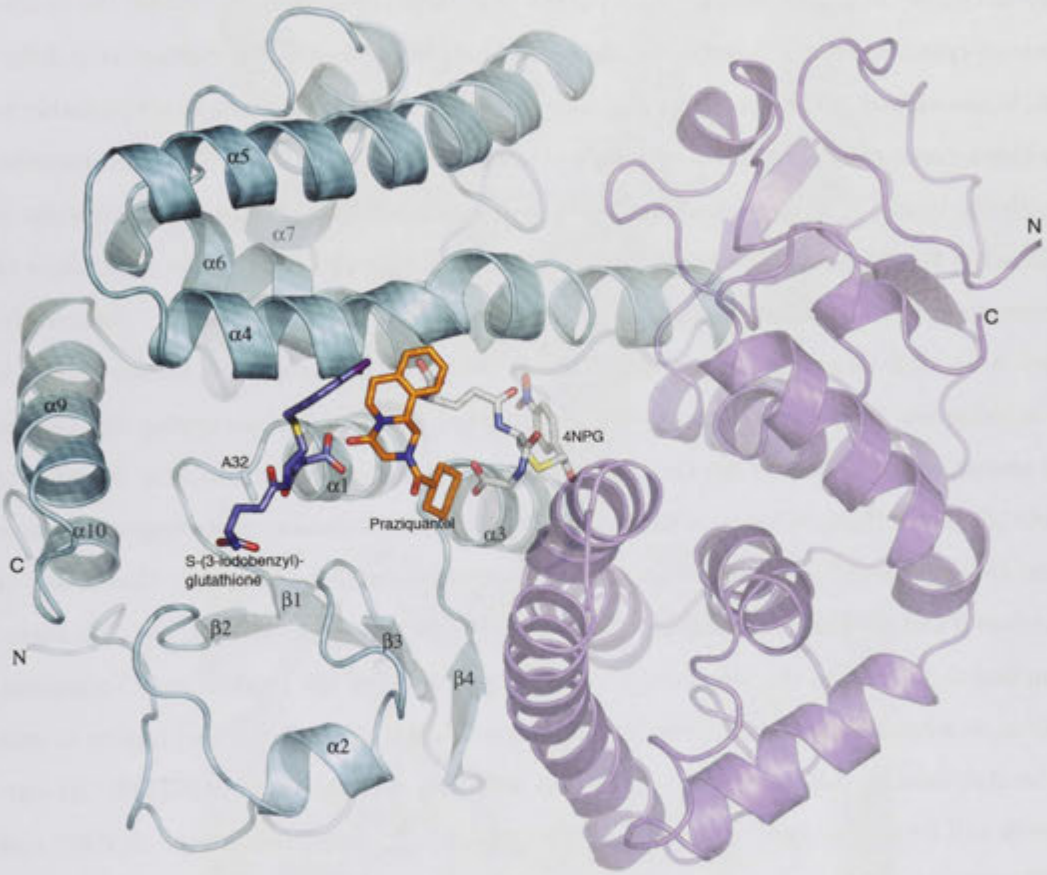


Figure 5.9

The crystallographic fold of hGSTO1 is shown as a cyan cartoon with its symmetrically generated subunit in magenta. The view is from above the physiological dimer, tilted to the left to allow visualisation of ligand binding at the interface. The binding mode of 4NPG has been compared to the 'ligandin' binding modes of S-(3-iodobenzyl)glutathione (orange) and Praziquantel (blue) in the σ class of squid and the α class of of *S. japonica* respectively. The crystal structures of these enzymatic complexes were superimposed onto that of hGSTO1 using the SSM algorithm in Coot [43]. The σ class S-(3-iodobenzyl)glutathione complex (PDB identifier: 2GSQ) superimposed with a rmsd value of 2.44 Å over 172 residues while the α class Praziquantel complex (PDB identifier: 1GTB) superimposed with a rmsd of 2.61 Å over 177 residues (not shown for clarity).

second is a Sigma class GST derived from the parasitic fluke *Fasciola hepatica*, FhGSTS1-1 (PDB identifier: 2WDU, unpublished). These have been overlaid with the hGSTO1 C32A GSSG/4NPG ternary complex presented here for comparison. Representations of both the monomers (Figure 5.10 on the facing page) and physiological dimers (Figure 5.11 on page 116) are shown. Inspection of GSSG binding within the former reveals that while the enzymes share analogous binding modes of the glutathionyl functionality within the G-site, the conformational space sampled by the second half of the disulfide is distinctively different. When viewed in context of the physiological dimers, this appears to be attributable to the characteristically open dimer interface of hGSTO1-1. Within FhGSTS1-1 for instance, the glycinyl motif of GSSG is stabilised by a hydrogen bond to N131 from the equivalent of helix α5 in the opposing monomer. Within hGSTO1-1 however, the relative orientation of monomers renders such an interaction impossible (Figure 5.11 on page 116). Conversely, there is no such interaction within hGSTM2-2. Although like most GST classes, there is an interdimeric interaction between the glutathionyl functionality occupying the G-site, the absence of these within the Omega class was previously noted not to alter its binding mode [4]. The difference between the human Mu and Omega classes rather appears to have arose from the evolution of a novel C-terminal extension within the latter. Comprised of α-helices 9 and 10, this has resulted in a 'deeper' H-site in the Omega class [4]. The corresponding region within the Mu class is occluded by residues of the random coil C-terminus, leading to a fundamentally different binding mode. Principally, the glycinyl moiety is seen to be stabilised by several polar interactions with the side chains of hGSTM2. Its carboxylic tail forms two salt bridge interactions with the guanidinium groups of R107 and R165 while its amide carbonyl is hydrogen bonded to the Y115 hydroxyl. The γ-glutamyl functionality on the other hand is orientated into the interdimeric cleft. There is a stark absence of such polar interactions with the H-site residues of hGSTO1. This, coupled to the relatively much deeper H-site pocket has results in a much 'tighter' mode of binding. Both ends of the H-site bound glutathionyl functionality are actually orientated away from the interdimeric cleft, rather than towards it. The evolution of a relatively open dimer interface that has facilitated the novel binding of 4NPG observed here, appears to have been coupled with a new recognition mode of GSSG. This is likely to have profound implications for this products binding affinity and subsequently, for the catalytic efficiency of all hGSTO1-1 mediated chemistry.

Figure 5.10
The structural fold of the hGSTO1 C32A monomer is shown as a cyan cartoon complexed with GSSG and 4NPG in stick representation, coloured according to atom type. Overlaid in translucent is the GSSG complexes of hGSTM2 (magenta) and FhGSTS (yellow). A magnified view of GSSG binding is also given for clarity. While binding of the ligands glutathionyl functionality is seen to be generally conserved within the G-site, the second half of the disulfide displays significant conformational heterogeneity.

Figure 5.11
The relatively open configuration of the hGSTO1-1 dimer in complex with GSSG and 4NPG is, contrasted to the GSSG complexes of the more closed hGSTM2-2 and FhGSTS1-1 dimers. The solvent inaccessible area at the hGSTO1-1 dimer interface with ligands removed is 1948 $Å^2$, compared to 2782 and 2637 $Å^2$ for the hGSTM2-2 and the FhGSTS1-1 dimers, respectively, (calculated using the SurfRace algorithm [110]). The physiological dimers above have been aligned to the structural unit on the left, accentuating the differences in the relative orientation of monomers. Colouring is as stated in Figure 5.10 on the preceding page. A close up of the GSSG binding modes has again been shown for clarity.

5.3.4 Comparisons with other GSTO orthologs catalysing the reduction of aromatic xenobiotics by deglutathionylation

The white rot fungus *Phanerochaete chrysosporium*, has recently been characterised to express 8 different enzymes that appear to belong to the GST Omega class, named PcGSTO1 to PcGSTO8 [111]. These have been categorized into 3 subclasses based on phylogenetic analysis and substrate specificity. The subclass II and III enzymes are the most closely related to mammalian orthologs such as hGSTO1, while the type I enzymes appear to be more closely related to bacterial GSTs [112] and the Lambda GSTs of plants [56]. The functional divergence of the Omega class within *P. chrysosporium* is reflected within the other GST families expressed by the organism, that display similar diversity. Combined with an equally diverse complement of phase I detoxifying Cytochrome P450 monooxgenases, the fungus is able to degrade many recalcitrant compounds including pesticides, halogenated aromatics, polyaromatics and textile dyes [113]. In addition, *P. chrysosporium* is the only known organism capable of completely degrading lignin to carbon dioxide and water.

Two representatives of the type 1 and type II sub-classes, PcGSTO1 and PcGSTO3 respectively, have recently been biochemically characterised by Meux *et.al.*, accompanied by the crystal structure of the former [113]. Both enzymes display characteristic thioltransferase and dehydroascorbate reductase activity that is dependent on an active site cysteine, however PcGSTO1 is also capable of catalyzing the reduction of S-glutathionyl-*p*-hydroquinone substrates, whereas PcGSTO3 is not. Conversely, only PcGSTO3 is catalytically active to 4NPG, and in fact displays a catalytic efficiency over an order of magnitude higher than observed for hGSTO1 [113]. Of particular interest, both native PcGSTO1 and PcGSTO3 are inactive towards CDNB, however, the glutathione conjugation activity of PcGSTO3 can be restored in analogous manner to hGSTO1, by mutating the active site cysteine to an alanine. The equivalent active site mutant of PcGSTO1 however, remains inactive [113]. This difference in enzymatic activity may be reconciled by comparing sequence conservation within their active sites. While PcGSTO3 contains a conserved FCPFVQ G-site sequence quite similar to the FCPFAE signature of hGSTO1, PcGSTO1 contains a CPWATR motif that is conserved in orthologs within other species, such as the three GSTO homologues identified in *S. cerevisiae*, designated Gto1 to Gto3 [114], and in bacterium such as *Sphingobium chlorophenolicum* and *E. coli*, named PcpF and YqjG respectively (Figure 5.18 on page 126). These enzymes have also been identified to function as S-glutathionyl-*p*-hydroquinone reductases [112]. This has prompted the suggestion

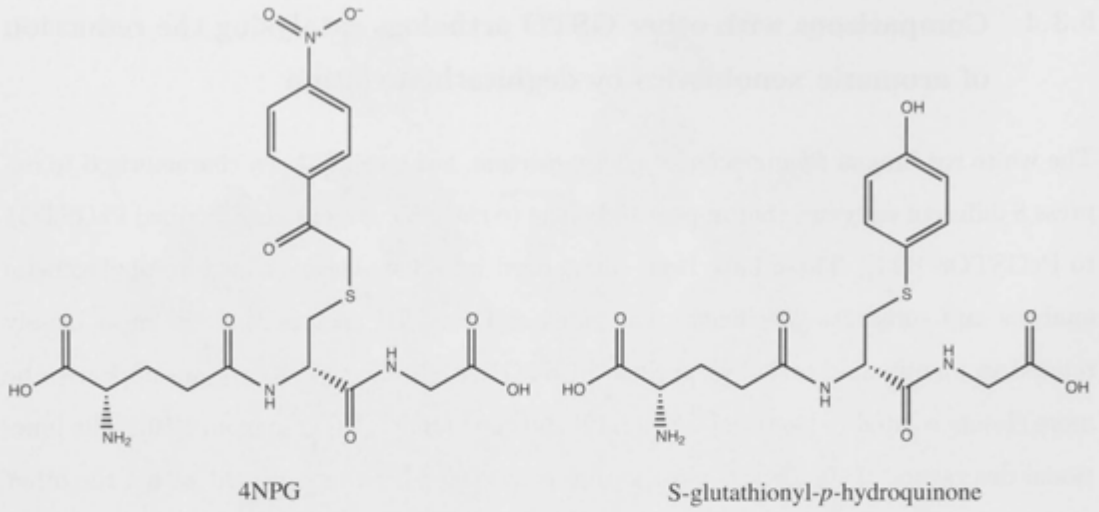


Figure 5.12
A comparison of 4NPG (left) to the structure of a typical S-glutahionyl-*p*-hydroquinone (right) substrate of PcGSTO1 [113].

for these enzymes to be classified within their own class, designated ξ (Xi) [112, 113]. Indeed, this is corroborated by the recent solution of the PcGSTO1-1 crystal structure. The enzyme displays a unique N-terminal extension of 77 residues, over 3 times the length observed in hGSTO1, and also possesses the novel α9 helix at its C-terminus. Rather than subsequently terminating with the short helix α10 however, the additional 20 residues of the PcGSTO1 C-terminal extension forms a coiled motif that also travels antiparallel to helix α9, but along its opposite face. This is seen to facilitate an entirely unique mode of dimerisation in the enzyme (Figure 5.13 on the facing page), which in contrast to all other GSTs, is exclusively mediated by the α-helical C-terminal domain. In addition, this C-terminal coil is prevented from adopting the same confirmation as hGSTO1 helix α10 due to this region of space being already filled by a 20 residue insert that forms an unusually long β2-α2 loop. [113]. In lieu of PcGSTO3, this structure (PDB identifier: 3PPU) has been overlaid with the hGSTO1/4NPG/GSSG ternary complex for comparison (Figure 5.13 on the next page). Despite the dramatically different mode of dimerisation, chain A of each superimposed with an rmsd value of 2.64 Å over 170 residues [43].

While the enzyme was observed in complex with GSH within only one of the 2 monomers within the asymmetric unit, the mode of ligand binding was observed to be highly similar to hGSTO1. The successive interactions of L71 and V72 with its glycinyl moiety are conserved, as are the polar interactions of its γ-glutamyl zwitterion with E85 and S86. Similarly, the hydrophobic interaction of this region with F34 has been conser-

Figure 5.13

The ternary complex of hGSTO1-1 with 4NPG and GSSG contrasted to the novel mode of dimerisation observed within the type I ortholog of *P. chrysosporium*. The crystallographic fold of hGSTO1-1 is shown as a cyan cartoon with that of PcGSTO1-1 overlaid in magenta. The novel N- and C-terminal extensions are highlighted in yellow and red respectively, while the 27 residue $\beta 2$-$\alpha 2$ loop is shown in black. Despite the disparity in their relative modes of dimerisation, equivalent chains aligned with an rmsd value of 2.64 Å over 170 residues [43]. The binding positions of the ligands 4NPG, GSSG and GSH are shown in stick representation, coloured according to atom type.

vatively replaced with an analogous rotamer of a tryptophan residue. Other interactions were nonetheless unprecedented. The guanidinium group of R56, belonging to the novel N-terminal extension of PcGSTO1, is seen to form a polar interaction with the cysteinyl-γ-glutamate carbonyl, effectively creating a 'lid' on the binding site. Similarly, the cysteinyl-glycine carbonyl accepts a hydrogen bond from the indole nitrogen of W119, located within the inserted β2-α2 loop structure. In addition, a slightly longer loop at the α2-β3 junction facilitates the formation of two polar interactions between the guanidinium group of R155 and the opposing glycinyl carboxylate of GSH (Figure 5.14 on the facing page). These interactions between the novel structural motifs of PcGSTO1 create a much more 'closed' active site relative to hGSTO1. The solvent accessible surface area of GSH being only 45 $Å^2$, compared to 145 $Å^2$ within the equivalent complex of hGSTO1 (Figure 5.14 on the next page) [4, 113]. Indeed, within the other opposing monomer in which GSH was not observed, residues 1-54 were disordered, suggesting that conformational change in the N-terminal extension may be required for the substrate to gain access to the G-site.

The most surprising feature of GSH binding to PcGSTO1 however, is perhaps the absence of a mixed disulfide bond with the active site residue C86. Its sulfhydryl was crystallographically observed to reside 3.9 Å away from that of GSH. The latter is rather seen to engage in a polar interaction with the side chain hydroxyl of Y223, itself participating in a tight hydrogen bonding network with nearby Y281 and Y326. This triad of tyrosine residues completely occludes the H-site of overlaid hGSTO1, suggesting a dramatically altered mode of GSSG binding. Indeed, while the hydrophobic nitro-benzene group of 4NPG can easily be modeled within this region of hGSTO1 (Figure 5.15 on page 122), the putative binding of S-glutathionyl-*p*-hydroquinone to PcGSTO1 would necessitate novel interactions whose precise nature are still unclear [113] (Figure 5.16 on page 123).

The 'ligandin' binding mode observed for 4NPG however, could be conserved. Although within the ternary complex of hGSTO1, 4NPG was observed along a crystallographic two fold corresponding to the dimer interface, the novel dimerisation of PcGSTO1 subunits has exposed the equivalent region to solvent. As could be expected, there has been a concerted shift away from the characteristically hydrophobic environment associated with this region of hGSTO1-1, to accommodate a complement of charged and polar side chains instead. Comparison of 4NPG binding with the overlaid structure of PcGSTO1 suggests this has not necessarily precluded an equivalent binding mode of S-glutathionyl-*p*-hydroquinone however (Figure 5.17 on page 125). The two residues observed to interact with both

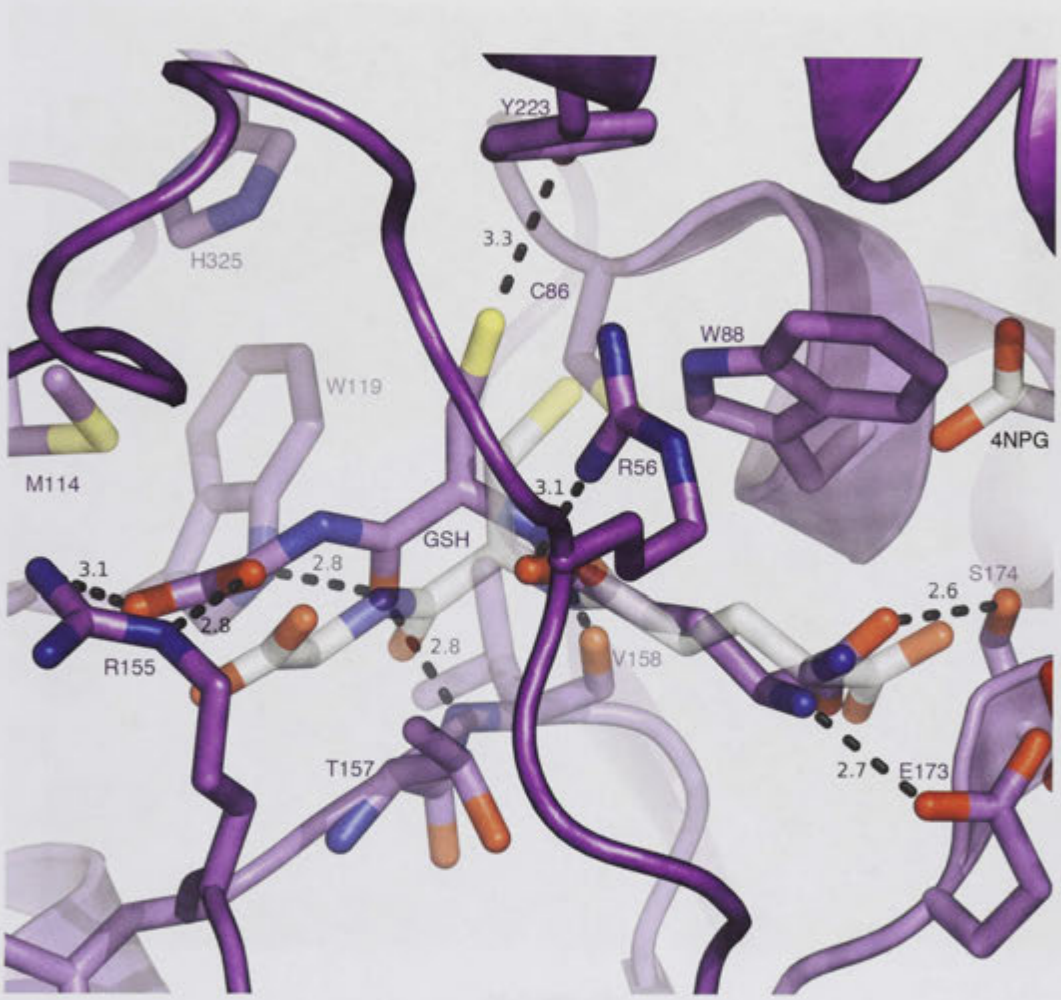


Figure 5.14

The G-site of PcGSTO1 is shown as a magenta cartoon with GSH and associated amino acid residues shown in stick representation. The binding of GSH, observed with an occupancy of 0.3 in the overlaid hGSTO1 complex with GSSG and 4NPG, is shown in translucent stick representation along with the nearby binding of 4NPG. GSSG is not shown for clarity. Polar interactions are shown as dashed lines and the crystallographically observed distances of each quoted in Å.

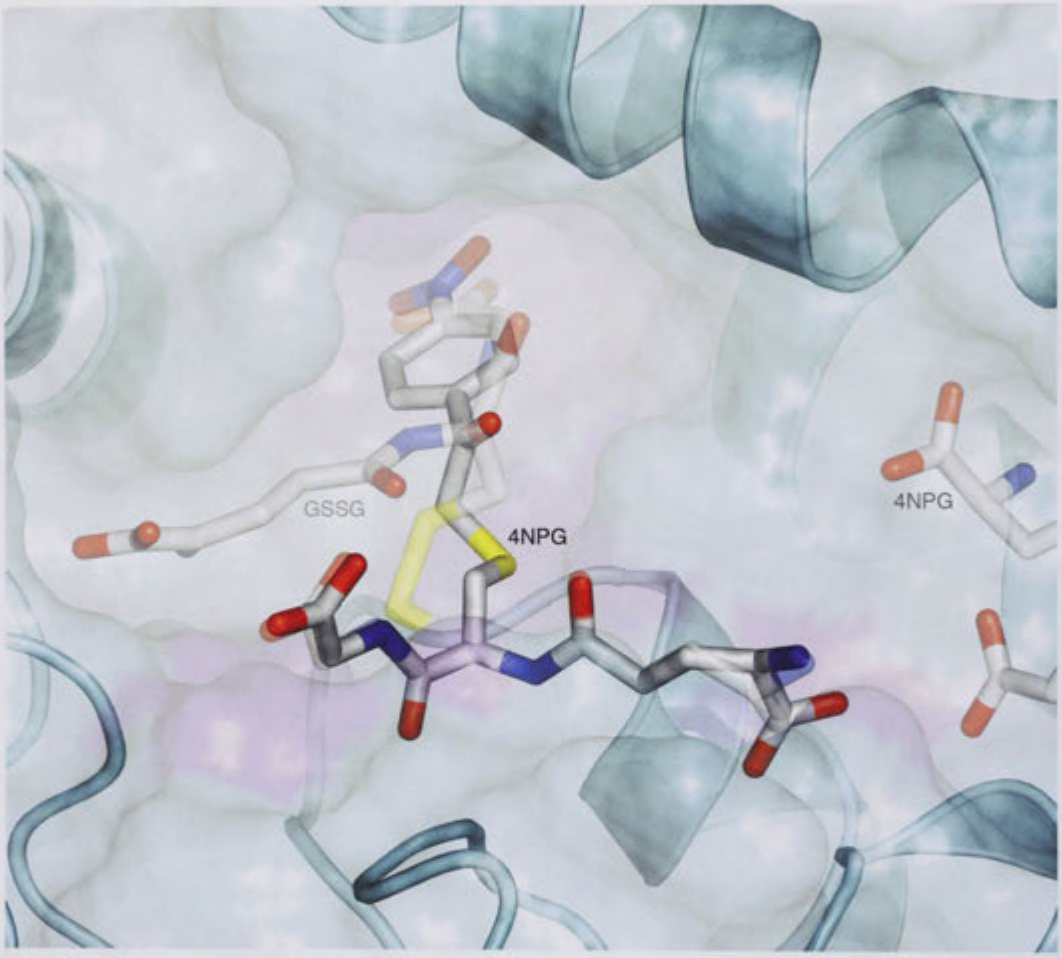


Figure 5.15

The enzymatic binding mode of 4NPG within the hGSTO1 H-site was modeled via real space refinement [43] into the $2mF_O - DF_C$ electron density corresponding to bound GSSG. Ligand binding is shown in stick representation, coloured according to atom type. The experimentally determined binding of GSSG and 4NPG are overlaid in translucent. The crystallographic fold of hGSTO1 is shown as a cyan cartoon, covered in a translucent molecular surface. The hypothesised binding mode of 4NPG in the active site is highlighted in magenta. Superposition of the PcGSTO1 crystal structure shows that an analogous binding mode of ligands would be impossible (Figure 5.16 on the next page).

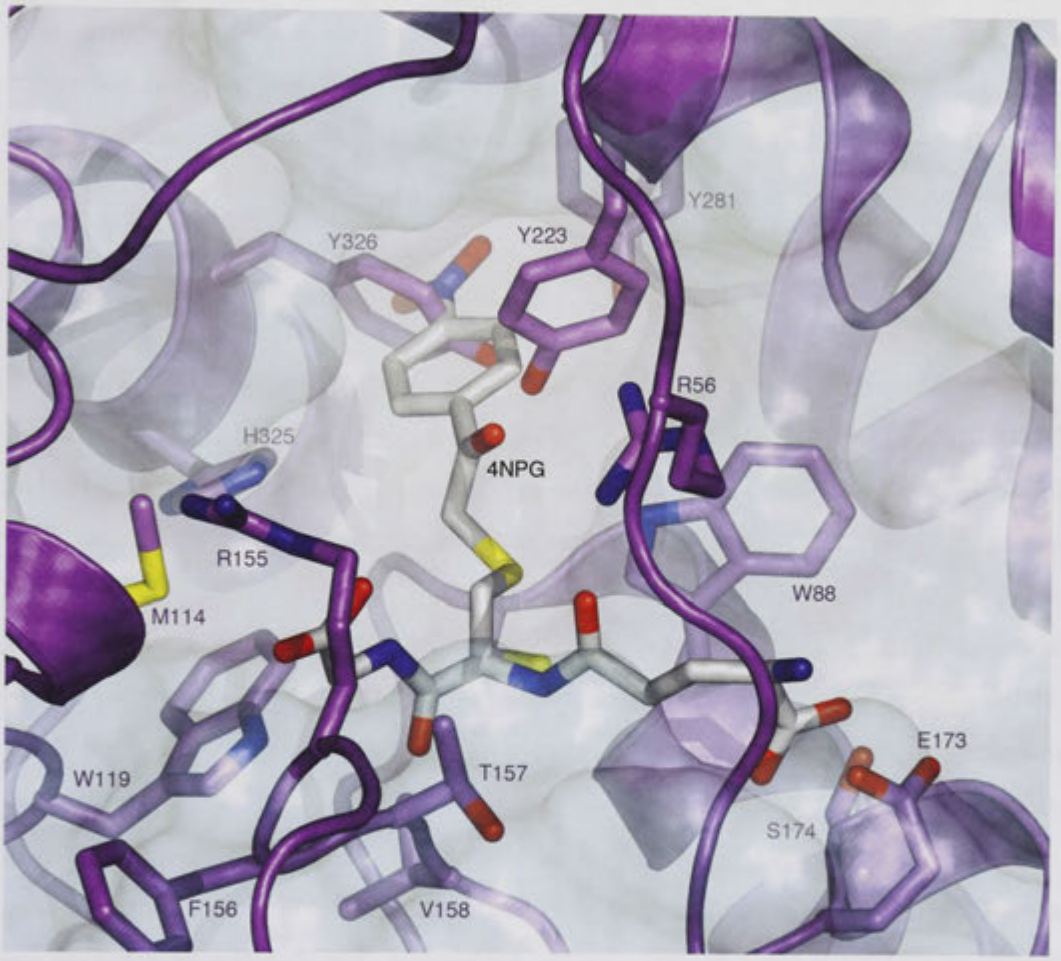


Figure 5.16
The superposition of 4NPG binding modeled in Figure 5.15 on the preceding page with the crystal structure of PcGSTO1 reveals that the H-site of hGSTO1 has not been conserved. Its occlusion by the side chains of 3 tyrosine residues, engaged in hydrogen bonding network to the thiol of GSH (not shown for clarity), suggests that the productive binding of glutathionylated aromatics and GSSG would feature novel interactions that are as yet uncharacterised. The translucent molecular surface of hGSTO1 is shown in cyan, with putatively bound 4NPG in stick representation, coloured according to atom type. The crystallographic fold and associated residues of PcGSTO1 are shown in magenta.

substrates within the hGSTO1 ternary complex of GSSG and 4NPG, R37 and E85, are both tightly conserved in the form of PcGSTO1 residues R91 and E173, respectively. While the sum of these substitutions would create a significantly different electrostatic environment for the binding of S-glutathionyl-*p*-hydroquinone, there is no obvious steric inhibition of it doing so. Of particular interest, the substitutions within this region of PcGSTO1 are, almost without exception, tightly conserved within the sequences of homologues from other species also suspected to belong to this subclass of Omega (GST Xi class), but not within the type II subclass more closely related to the mammalian enzymes, expected to have the same dimerisation mode as hGSTO1-1 (Figure 5.18 on page 126).

5.3.5 Structural basis for substrate specificity within the Omega class

A question that must be asked about the phenacyl glutathione reductase activity of hGSTO1-1, is why it is not shared by its homologous isozyme, hGSTO2-2? The recent solution of the hGSTO2 crystal structure by my colleague, Dr. Huina Zhou (unpublished data), allows a detailed structural argument to be framed in response. Structural alignment of the isozyme with the hGSTO1 ternary complex is achieved with a rmsd value of 0.94 Å over 238 residues [43], and the region of the dimer interface associated with binding of 4NPG is observed to be remarkably conserved (not shown). The only changes within the immediate vicinity of the substrate are the conservative replacements of L103 and K110, by F103 and R110.

The disparity in enzymatic activity is therefore likely to arise from chemical differences associated with the putative binding of 4NPG within the active site. Within this region, there are several inconsistencies that may be of relevance. The structure of hGSTO2 was solved in complex with GSH in the G-site, and its mode of binding, as well as the associated residues, are highly similar to those of hGSTO1. The non-conservative L71H substitution noted previously (Figure 3.10 on page 59) is unlikely to affect binding of 4NPG, since it would be expected to interact with its glutathionyl functionality in a similar confirmation. The significant structural heterogeneity within the H-site is a much better candidate for the observed differences in activity (Figure 5.19 on page 128 and 5.20 on page 129). Based on similar complexes of GSTs with derivatives of 1-glutathionyl-2,4-dinitrobenzene [33, 115, 116, 117, 6], the nitrophenacyl moiety of 4NPG would be expected to bind in the manner shown in Figure 5.15 on page 122. Within hGSTO2 however, several non-conservative amino acid substitutions would make such a binding mode difficult. Perhaps the most

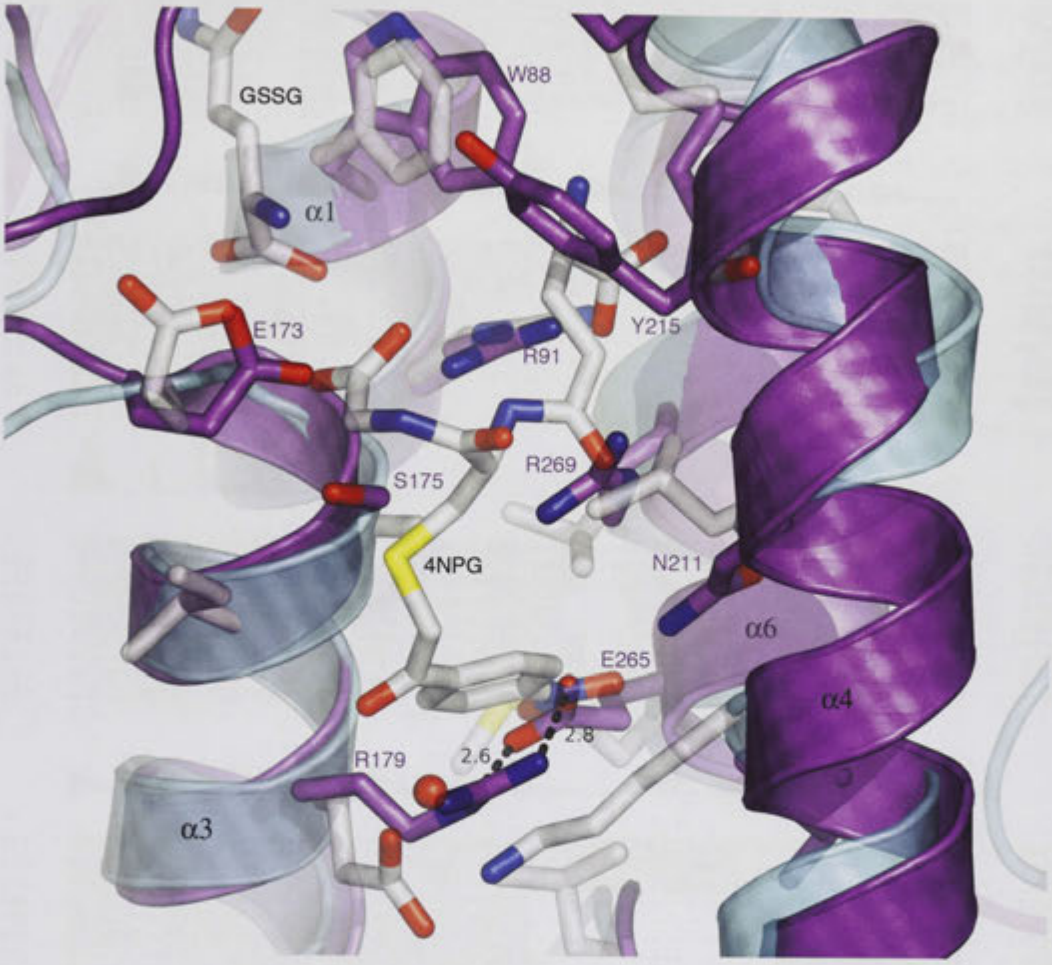


Figure 5.17

A superposition of the 4NPG binding mode to hGSTO1 with the equivalent region of PcGSTO1 suggests it may be transposable to the binding of S-glutathionyl-*p*-hydroquinone. Despite several non-conservative amino acid substitutions within the *P. chrysosporium* enzyme, only minor conformational changes in the associated amino acids would be required to accommodate an equivalent binding mode. The structural fold of hGSTO1 is shown as a cyan cartoon, with the observed binding of 4NPG and associated side chains shown in translucent stick representation, coloured according to atom type. The equivalent region of the PcGSTO1 is overlaid in magenta, which due to its novel mode of dimerisation, is now located at the solvent interface. Putative residues associated with an equivalent binding mode of S-glutathionyl-*p*-hydroquinone are labelled accordingly. The equivalent residue to E91 within hGSTO1-1, that forms an intermolecular salt bridge with K114 in the opposing monomer, has been replaced by R179 within PcGSTO1, and now forms an equivalent intramolecular interaction with residue E265 from nearby α6. This interaction is indicated with dashed lines, the crystallographically observed distances of each is quoted in Å.

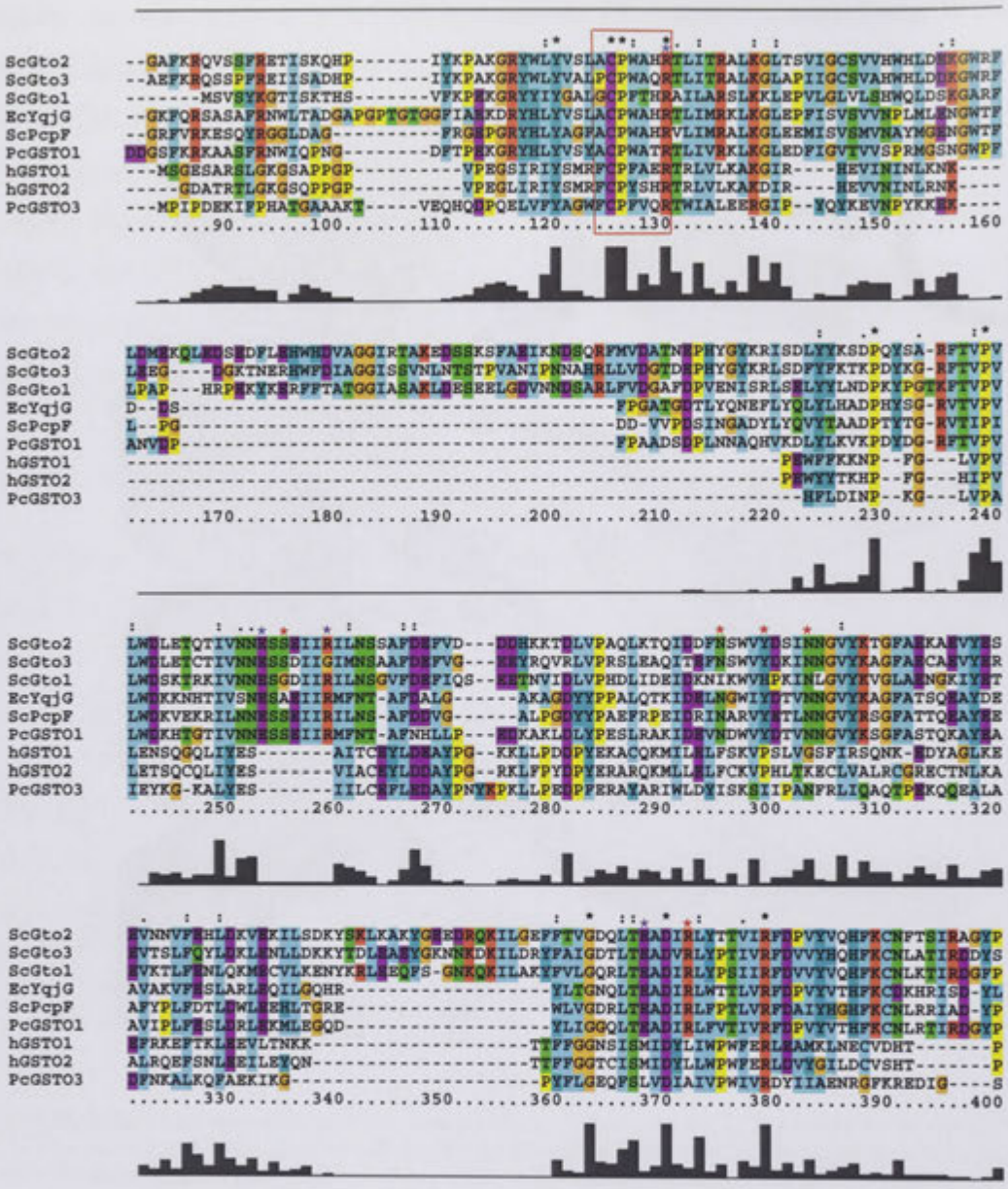


Figure 5.18

The multiple sequence alignment of PcGSTO1 with members of the GST Xi class from yeast (ScGto1-3), *E. coli* (EcYqjG) and *S. chlorophenolicum* (ScPcpF), compared to the the human Omega class enzymes and PcGSTO3. The highly conserved active site motif is outlined in a red box. Based on the complex of hGSTO1 with 4NPG, residues hypothesised to feature in a 'ligandin' binding mode of S-glutathionyl-*p*-hydroquinone are marked with coloured asterisks. Those also associated with binding of GSH are marked with a blue asterisk, while the positions of the R179-E256 salt bridge residues are marked in magenta. The rest are marked in red. Almost without exception, these are tightly conserved within the putative Xi class enzymes, but not within PcGSTO3, expected to possess a mode of dimerisation analogous to the human Omega class. (NB: The residue numbering of PcGSTO1 in this figure is positive by a value of 31 with respect to the rest of the text. This is due to the sequence containing a signaling peptide that was excised prior to purification of the recombinant enzyme [113]).

striking is the substitution of M187 for Y188, which in a manner reminiscent of residue Y223 from PcGSTO1 (Figure 5.16 on page 123), is observed orientated downwards, partially occluding the H-site. Intriguingly, this residue is located at the start of the α6-α7 junction, and may also affect enzymatic activity via its steric bulk restricting the conformational change associated with substrate binding to hGSTO1 (Figure 5.5 on page 108). In addition, a radical replacement of G128 with a lysine has occurred at the entrance to the hGSTO2 H-site. Although 5.3 Å away from the hydroxyl of Y230 from the shifted helix α9, these two residues effectively form a 'polar gate' , that together with the side chain of Y188, may make the binding of hydrophobic moieties in this region electrostatically and sterically impossible. It is possible that a previous gene duplication of an ancestral *hGSTO1* sequence may have allowed divergent evolution away from its structural features adapted for the detoxification of hydrophobic xenobiotics, subsequently facilitating the development of this enzymes exceptional efficiency towards the reduction of dehydroascorbate.

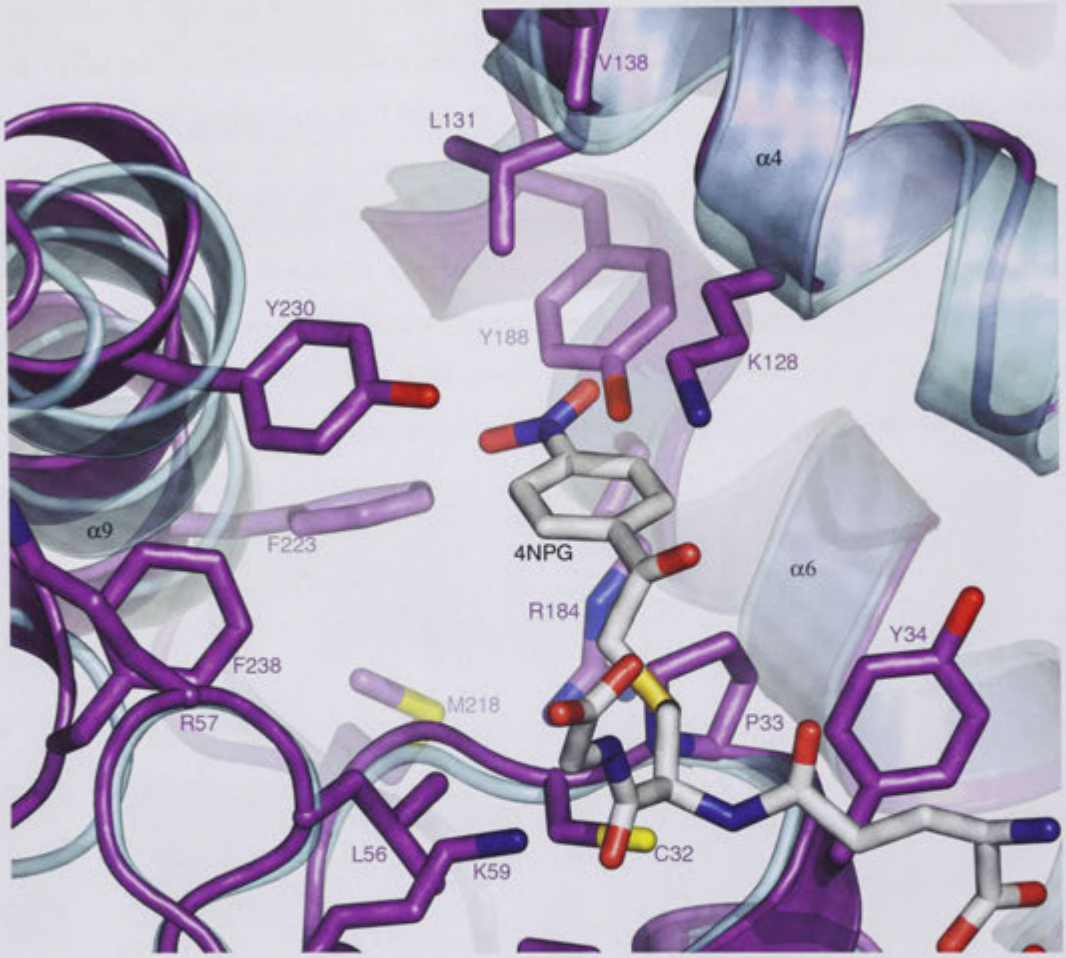


Figure 5.19

The superposition of hGSTO2 (Dr. Huina Zhou, unpublished data) H-site residues with the modeled binding of 4NPG to hGSTO1 shown in Figure 5.15 on page 122 reveals several immoderate substitutions that may contribute to their differing enzymatic activity towards S-phenacyl glutathione substrates. The side chains of hGSTO2 (magenta) show that the replacement of a methionine by the side chain of Y188 results in a significant occlusion of the H-site pocket. In addition, a glycine of helix α4 has been replaced by a lysine at position 128, significantly decreasing the size and hydrophobicity of this enzymatic cleft. The structural fold of the hGSTO1 ternary complex is shown as a cyan cartoon with the putative binding of 4NPG shown in stick representation, coloured according to atom type.

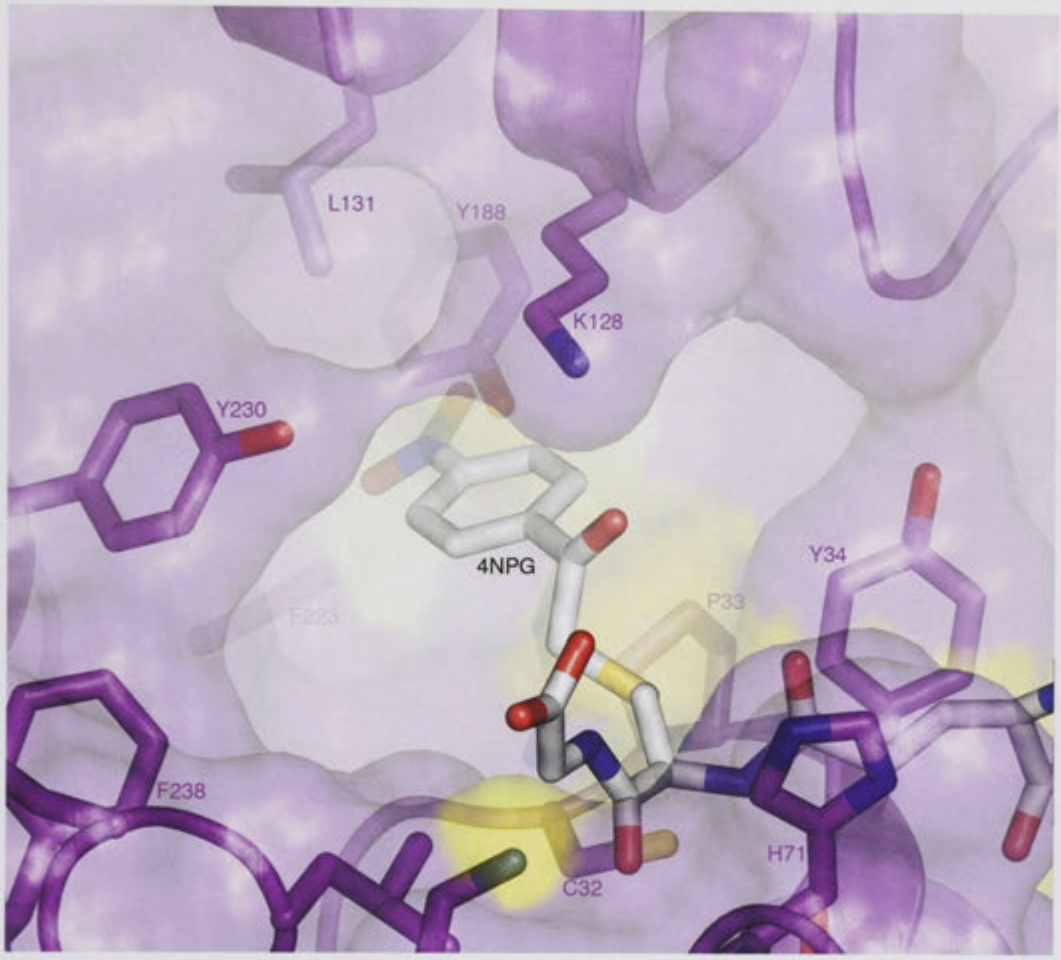


Figure 5.20

A translucent molecular surface representation of the hGSTO2 active site relative to the putative binding mode of 4NPG binding within the equivalent region of hGSTO1 shows significant occlusion of the H-site. The structural fold, molecular surface and relevant side chains of hGSTO2 are shown in magenta, with hypothetical contacts of 4NPG highlighted in yellow. Of most interest are the residues K128 and Y188. While the hydroxyl group of the latter resides in a position that clashes with the *para*-nitro group of 4NPG, the non-conservative replacement of an α4 glycine with a lysine has resulted in dramatic size reduction of this binding cavity. Together with the opposing side chain of Y230, this has effectively formed a 'polar gate' to the H-site. While a narrow channel is still present below these residues, it is most likely too narrow and electrophilic to accommodate the bulky phenacyl moieties of substrates such as 4NPG. I am indebted to Dr. Huina Zhou for allowing me to use the coordinates of hGSTO2 prior to publication.

Chapter 6

Efforts to discover the mode of hGSTO1-1 inhibition

A limited additive screen and altered purification conditions has led to the discovery of two novel crystal forms

Alzheimer's disease is a neurological disorder that has long been suspected to involve the aberrant expression of the pro-inflammatory mediator IL-1 as a causative factor [118]. The discovery that hGSTO1-1 was the target of a class of diarylsulfonyl urea drugs known to inhibit the post translational modification of this cytokine was therefore provocative, especially given the previous associations made between this enzyme and the age at onset of Alzheimer's disease [119].

Yet another type of hGSTO1 inhibitor has recently been discovered during the testing of a high-throughput method of selective inhibitor discovery involving a fluorescent activity-based probe [32]. Due to upregulation of hGSTO1-1 in several cancer cell lines that show enhanced aggressiveness and therapeutic resistance, it was a prime candidate for these researchers to trial their new method. This resulted in the identification of an α-chloroacetamide polyphenol that could potently and selectively inhibit the enzyme. An aim of this research was to characterise the mechanism by which these compounds inhibited hGSTO1 crystallographically. While these efforts to date have been unsuccessful, they have identified two novel crystal forms of the enzyme. The subsequently solved crystal structures have provided valuable insight into the structural dynamics of hGSTO1 and are discussed in more detail below.

6.1 Introduction

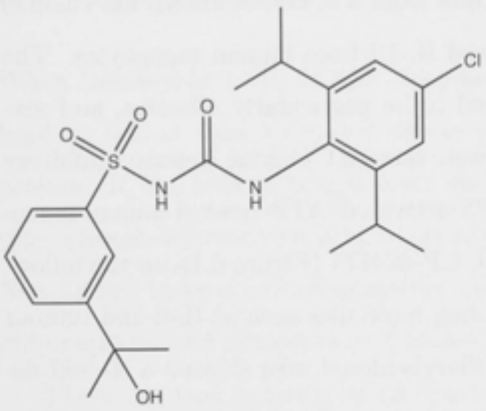
When Interleukin 1 (IL-1) was discovered in the late 1970's, many researchers found it hard to believe that its varied effects within the body could be attributed to just one protein. It was able to trigger fever and activate white blood cells, but also appeared to play a tissue destructive role in arthritic joints and influence other proteins within the liver. Now known to be a pro-inflammatory cytokine, it has been implicated in a wide variety of inflammatory and autoimmune diseases [120].

There are two isoforms of the protein, encoded by related but separate genes, and their transcribed products are known as IL-1α and Il-1β, respectively [121]. The latter is primarily produced by activated cells of the monocytic lineage, including monocytes, macrophages and dendritic cells. While this cytokine has an essential and beneficiary role in mediating responses to microbial attack, it can also be damaging. The protein has been associated with an increase in tumour malignancy and appears to facilitate tissue damage in the pathogenesis of chronic auto-inflammatory conditions such as arthritis [122]. The cytokine is therefore not constitutively expressed, but is transcriptionally upregulated in response to a variety of stimuli. These typically fall into two categories; (i) Those associated with the invasion of pathogens, such as lipopolysaccharide (LPS), and (ii) those associated with cell damage, such as extracellular ATP [122, 123]. Unlike most other cytokines however, IL-1 requires post-translational processing from its procytokine form in order to be efficiently exported from the cell and invoke a response. The post-translational processing of IL-1β is known to be mediated by the enzyme caspase-1, that cleaves the 31 kDa translation product into a mature 17 kDa cytokine. This exposes an N-terminal signal sequence that allows it to be efficiently exported from the cell [122]. Caspase-1 is now known to be regulated by large intracellular protein complexes known as inflammasomes. While the precise mechanism by which this occurs is still unclear, the efflux of K^+ ions from the cell appears to be a key upstream trigger of their activation [122]. In response to extracellular ATP for instance, the $P2X_7$ receptor initiates a cascade of chemical events that eventually lead to increased IL-1β post translational processing and release. Immediately following the binding of the nucleotide, this receptor undergoes conformational change to function as a non-selective cation channel. This is also known to activate several stress kinases, transcription factors, phospholipases, cause membrane depolarization and in some instances, apoptosis [122, 24].

Based on the premise that ionic changes stimulate the post-translational modification

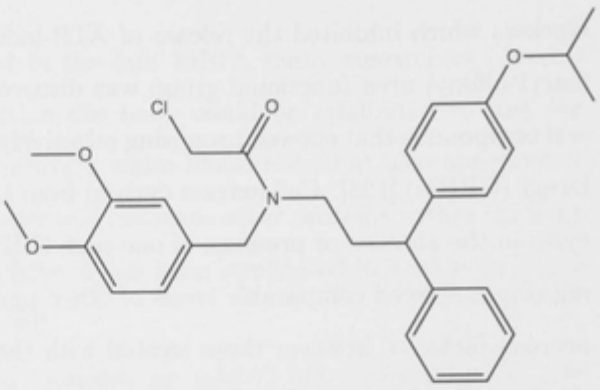
of IL-1, Pfizer has identified a series of lead molecules from a screen of known ion channel blockers which inhibited the release of ATP-induced IL-1β from human monocytes. The diarylsulfonyl urea functional group was discovered to be particularly effective, and several compounds that showed promising selectively were termed Cytokine Release Inhibitory Drugs (CRIDs) [123]. Cell extract derived from LPS-activated/ATP-treated human monocytes in the absence or presence of one such CRID, CP-424174 (Figure 6.1a on the following page), showed comparable levels of other signaling molecules such as IL-6 and tumour necrosis factor-α, however those treated with the diarylsulfonyl urea showed a 30-fold decrease in the levels of Il-1β [123, 121]. While the precise mode of action was unknown, they were known not to be direct inhibitors of caspase-1 or the interaction of ATP with $P2X_7$. Traditional biochemical experiments were therefore performed to discover its target within the proteome. When extracts of human monocytes were passed through a CRID affinity column, hGSTO1-1 bound selectively and was eluted by a soluble form of the drug. In addition, hGSTO1-1 readily incorporated ^{14}C-labelled epoxide containing CRIDs in an irreversible manner. This was dependent on the presence of an active site cysteine and correlated with an inhibition of cytokine production. Interestingly, a CRID-glutathione adduct was also capable of inhibiting the ATP-induced post-translational processing of the cytokine [24].

These findings are provocative given the region of chromosome 10q containing *hGSTO1* and *hGSTO2* has been associated with an earlier age at onset (AAO) of both Parkinson's and Alzheimer's disease in genetic mapping studies [119] and that these conditions are characterised by augmented levels of IL-1 expression [118, 122]. A SNP within exon 4 of *hGSTO1* is associated with a common A140D polymorphism that was subsequently found to associate with the AAO of both conditions, as was another in the promoter region of *hGSTO2*. The association of these polymorphisms could not be confirmed in other populations however [124, 125]. In addition, immunostaining of the atherosclerosis plaques of peoples brains suffering from the closely related condition of vascular dementia and carrying the A140D polymorphism revealed a co-localization of hGSTO1-1 and IL-1α [23]. Furthermore, IL-1 is differentially and abnormally over-expressed in the brains of Alzheimer's patients. A polymorphism within the promoter region of the IL-1α gene, and another within a coding region of IL-1β gene, have both been associated with the over-expression of the cytokine and infer an increased risk of developing several auto-immune conditions. Patients homozygous for the IL-1α polymorphism for instance, are up to 3



IC_{50}=210 nm

(a) An example of the diarylsulfonyl urea class of CRIDs found to target GSTO1-1 in human monocytes by Laliberte *et al.*, designated CP-424174 [24].



IC_{50} = 120 nm

(b) The potent and selective α-chloroacetamide inhibitor of hGSTO1-1 discovered by Bachovchin *et al.*, [32].

Figure 6.1
Two selective and potent inhibitors of hGSTO1-1.

times more likely to develop Alzheimer's disease, and are likely to do so approximately 10 years earlier than average. If patients are homozygous for both polymorphisms, the risk is compounded to over 10 times greater than the rest of the population [118].

The inhibition of the Omega class has also been pursued directly. The fact that hGSTO1-1 is over expressed in several human cancer cell lines that show enhanced aggressiveness and chemotherapeutic resistance made it an ideal model target for a recently developed method of selective inhibitor identification using high-throughput screening with fluorescent activity-based probes. Using this technology, these researchers were able to identify a polyphenol that was a potent inhibitor of the enzyme. It was determined that this α-chloroacetamide possessed an IC_{50} = 120 nM (Figure 6.1b), using a 4NPG assay [32].

During this research, samples of both types of inhibitor were obtained from their manufacturers. The aim of subsequent experiments was to crystallographically characterise their mode of hGSTO1-1 inhibition. To date, these efforts have not been successful, but have resulted in the discovery of two novel crystal forms of hGSTO1, whose structures have been solved at resolutions of 3.21 and 1.81 Å, respectively. The different methodologies employed, possible reasons for their failings and the insights into hGSTO1 structural dynamics afforded by two novel crystal lattices are discussed below.

6.2 Experimental Methods

One of the major hurdles facing any crystallographic investigation of protein interaction with lipophilic compounds is the inherent insolubility of the latter under the saturating conditions associated with crystal formation. The known crytsallisation conditions of hGSTO1-1 for instance [4], features a precipitating agent of high salt (2.2 M $(NH_4)_2SO_4$), creating a fundamental difficulty in providing the relevant inhibitors (Figure 6.1 on the facing page) at the concentration required for observation crystallographically. Another that was faced during this research was the apparent redox sensitivity of the native enzyme. Repeated attempts at observing C32-thiol adducts with substrates and inhibitors within crystals of the native enzyme were consistently unsuccessful despite the controlled variation of other variables, such as inhibitor concentration, the presence of glutathione and pH. It was hypothesised that this may have been due to the oxidation of this residue to the sulfenic, sulfinic or sulfonic acid derivatives, or the inability to reduce the mixed disulfide bond between the active site cysteine and GSH once it was formed. Such problems in the ArsC enzyme of *S. aureus* (Section 4.3.4 on page 85) have previously been noted in the literature [96, 93, 94]. Crystallographic observation of the active site cysteine-arsenical adduct required careful secondary purification with reverse phase HPLC, using buffers previously deoxygenated with argon gas, in order to purify and crystallise the catalytically active enzyme. The present investigation was further complicated by the use of the reducing agent DTT within the published methods of hGSTO1-1 purification [4]. Since the selective inhibitors of hGSTO1-1 shown above (Figure 6.1 on the facing page) contain thiol reactive functional groups, it was possible that chemically formed adducts with DTT were also inhibiting their observation crystallographically.

To address these issues, several strategies were adopted from a comprehensive review of the literature on this topic [126] and combined with strategies applied to the *S. aureus* enzyme [96, 93, 94]. One strategy involved incubating the enzyme with its inhibitors before it was concentrated. This approach had two advantages. By introducing the inhibitor prior to crytsallisation, it avoided the inherent insolubility associated with the high salt concentration. Although still only marginally soluble in the aqueous protein/buffer solution, the low concentration of the enzyme allowed a much higher inhibitor to enzyme molar ratio to be achieved.

In addition, the 5 mM DTT included within the final dialysis buffer of previous hGSTO1-1 purifications was replaced with a 1 mM concentration of its physiological co-

substrate GSH instead. While the monothiol of GSH is a less effective reducing agent, its use in combination with careful deoxygenation of buffers with argon prior to use proved sufficient for maintaining protein solubility. It was hoped that this would prevent non-physiological reactions with ligands, that may have prevented the observation of their binding modes crystallographically. Unexpectedly however, subsequent hanging drop/vapour diffusion experiments using conditions analogous to those identified previously [4], resulted in the formation of crystals displaying a completely different morphology, subsequently found to belong to the tetragonal space group $P4_222$ (Figure 6.3 on page 141).

Finally, the review [126] had also noted success in observing complexes with lipophilic compounds via the use of limited additive screens, in which additive compounds involved in the crytsallisation of such complexes within the PDB were combined with known crytsallisation conditions of the enzyme. One of the screened compounds suggested was guanidine hydrochloride, and its addition to the crystallisation conditions has resulted in a transition from the previously observed trigonal crystal lattice to orthorhombic crystals of space group $C222_1$ (Figure 6.4 on page 144).

The experimental parameters associated with this altered methodology of protein purification, and the subsequent crystallisation of these novel crystal forms are detailed below.

6.2.1 A redox sensitive purification protocol of hGSTO1 native leads to two new crystal forms

A 50 μL aliquot of electroporation competent *E. coli* cells of strain BL21 were removed from storage at -80°C and transformed with 2 μL of *pHUE.hGSTO1* plasmid via electroporation. 0.5 ml of LB was then added immediately after the application of current to facilitate cell recovery. 100 μL of this solution was then spread onto a LB agar plate (0.2 mg/ml Ampicillin) using sterile technique, dried face up and then incubated face down overnight at 37°C. The next morning, multiple colonies were observed on the agar plate. One was used to inoculate a 50 mL aliquot of LB media (0.2 mg/ml Ampicillin) and subsequently incubated at 37°C overnight with gentle shaking. The next morning, bacterial proliferation was observed, and 10 mL aliquots of this starter culture were used to inoculate 5 respective 1 L volumes of media. This LB also contained 0.2 mg/ml of Ampicillin but in contrast to previous purifications, was also supplemented with 1 mM of GSH. These cultures were subsequently shaken at 37°C for ≈4 hours until their optical density values at 600 nm were in the range of 0.6-0.8, before inducing with 300 μL of 200 mM IPTG. After culturing

for a further 3.5 hours, the media was spun down in a 8K-10 centrifuge (Sigma) at 6500 rpm for 30 minutes, using a Nr.12510 rotor at 4°C. The supernatant was then discarded, and each of the 5 bacterial pellets resuspended in 50 mL of Buffer A (50 mM sodium phosphate, 300 m NaCl, 10% glycerol, 1mM GSH, 30 mM imidazole, pH 7.5) that had been flushed with argon gas for several minutes immediately prior to use and contained 0.1 mM of the protease inhibitor AEBSF. This was performed overnight using a mechanical stirrer bar at 4°C. The following day, the homogenous cellular resuspensions were lysed using a pre-chilled French pressure cell. The lysate was collected into centrifuge tubes and after removing floating nuclear material with a transfer pipette, spun down as above. The supernatant was then carefully decanted to exclude cellular debris, before vacuum filtering through 0.45 μm filter paper (Millipore) in the cold room. The filtrate was then loaded onto a 5 ml 'His-trap' Ni-NTA column (GE Healthcare) using a peristhetic pump. An FPLC mediated imidazole gradient was then applied to the column from 0 to 500 mM, derived from buffer A that had been deoxygenated with argon gas. Due to the relatively high amount of imidazole in the loading buffer (30 mM), non-specific binding to the column was observed to be minimal, and absorbance at 280 nm indicated that the vast majority of protein had eluted near the start of the gradient. Subsequent analysis of these fractions via SDS-PAGE revealed them to almost exclusively contain significant quantities of relatively pure hGSTO1-1. The most pure fractions were pooled, and their combined volume found to be $\approx$95 ml. A cleavage reaction was then performed via the addition of GSH and the Usp2 ubiquitinase to a final concentration of 10 mM and 25 μg/ml respectively, followed by incubation at 4°C overnight.

The protein solution was then dialysed against Buffer A (0 mM imidazole) that had previously been flushed with argon gas, and subsequently reloaded onto the re-equilibrated Ni-NTA column as above. The flow through was collected in a conical flask and the column was then washed with 3 CV of buffer A containing 0 mM imidazole, followed by a further 3 CV containing 500 mM of imidazole. Subsequent SDS-PAGE analysis revealed the cleavage reaction to have proceeded with near 100% efficiency. A high quantity of pure, deubiquitinated hGSTO1 was present in the flow through and initial column wash. The combined solutions were then dialysed in argon flushed final dialysis buffer overnight (20 mM Tris, 60 mM NaCl, 1 mM GSH, pH 8).

Approximately $^2/_3$ of the protein solution was then concentrated using a 10 kDa MWCO centrifugal filter device (Amicon). When the volume had been reduced to approximately

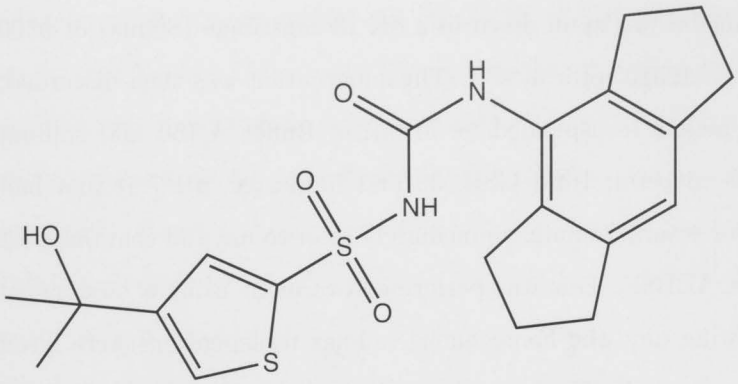


Figure 6.2
The diarylsulfonyl urea CP-457324.

1 ml, the protein concentration was assayed via absorbance at 280 nm using a 'ND-1000' spectrophotometer (NanoDrop Technologies), and found to be 17.23 mg/ml. As this was deemed a sufficient concentration to facilitate crytsallisation, the protein solution was stored at 4°C for later use.

The remaining 1/3 of the unconcentrated protein solution was then used for dilute inhibitor incubation experiments described in the literature [126]. This technique effectively allowed the protein to lipophilic inhibitor ratio to be increased to values that were impossible given the high salt environment of the published crytsallisation conditions [4]. After a period of incubation in which inhibitor binding and/or covalent modification could occur, the protein solution was subsequently concentrated and crytsallisation experiments performed. The former step was performed at room temperature in the hope the associated increase in reaction rate would be sufficient to ensure inhibition was homogenous throughout the solution [126]. In order to determine at what concentration the inhibitors should be combined with the protein solutions, their reported IC_{50} values were considered. In the recent fluorescent activity-based probe identification of the α-chloroacetamide inhibitor (Figure 6.1b on page 134), a 4NPG assay containing varying concentrations of the compound determined its IC_{50} value to be 120 nM [32], while the *in vivo* IC_{50} value of the diarylsulfonyl urea CP-424174 (Figure 6.1a on page 134) has been reported as 210 nM [127]. A sample of a related diarylsulfonyl urea, CP-457324 (Figure 6.2), whose structure and activity are unpublished, had also been obtained from Pfizer pharmaceuticals and similarly employed in these experiments.

Before the inhibitors were combined with purified hGSTO1-1, it was first decided to remove the 1 mM of GSH from the storage buffer by dialysis. This step aimed to min-

imise its capacity to compete with inhibitor binding in the G-site, and to allow formation of secondary reaction products with the thiol sensitive compounds. 5 L of final dialysis buffer was therefore prepared containing 20 mM Tris, 300 mM NaCl, pH 8.5, and flushed with argon gas for 10 minutes prior to use. After testing a small volume for precipitation, the protein solution was dialysed overnight at 4°C. This effectively diluted the amount of GSH present to 6.4 μM. The next morning the concentration of the protein solution was ascertained to be 0.55 mg/ml. Dividing by the calculated molecular weight of the enzyme, 27565.86 Daltons, the molar concentration was determined to be 20 μM. The compounds were found to be most readily soluble in DMSO (dimethyl sulfoxide), and a 1 mM solution was made of each. The 10.7 ml aliquots of protein were then diluted to a final volume of 28.2 ml with the final dialysis buffer described above (freshly flushed with argon gas) and combined with 300 μL of a respective 1 mM inhibitor solution. This resulted in the final concentrations of protein, inhibitor, GSH and DMSO being approximately equal to 7.6 μM, 10 μM, 2 μM and 1% (v/v), respectively. The solutions were subsequently incubated at room temperature for 2.5 hours before being stored on ice, and then concentrated using a respective 10 kDa MWCO centrifugal device (Amicon) at 4°C. When each aliquot was reduced to a final volume of between 100-150 μL, the hGSTO1-1/α-choloracetamide/CP-424174/CP-4573234 solutions were found to be at concentrations of 25.18, 15.21 and 16.50 mg/ml respectively. As these concentrations were deemed sufficient to facilitate crytsallisation, they were stored at 4°C and used to set up crytsallisation screens the following day.

6.2.1.1 Experimental conditions associated with hGSTO1 crytsallisation and structure solvation in a tetragonal crystal lattice

Strikingly, a precipitant versus pH crytsallisation screen at 4°C based on the previously published conditions [4] using the inhibitor/protein solutions described above resulted in crystals of a completely different morphology (Figure 6.3 on page 141) and were subsequently discovered to belong to the tetragonal space group $P4_222$ (Table 6.1 on page 141). Although the crystal structure presented here originated from a crystal grown from the 'dilute inhibitor incubation' protein preparation detailed above, this step was not essential for crystallisation, since protein concentrated without performing this step also crystallised in this space group. Thus, this unexpected change in crystal morphology appears to be due to the altered protocol of protein purification (Section 6.2.1 on page 136). Although

this crystal form presented the opportunity to crystallographically study the chemistry of hGSTO1-1 in the absence of non-physiological thiol reductants, crystals large enough for analysis by X-ray diffraction ($\geq$100x100x100 μm) were less common than the trigonal form studied previously. In light of this, and the limited supply of protein, efforts were made to increase the number of experimentally viable crystals via the technique of 'streak seeding'. This involved inspecting the drops of the crytsallisation screen after equilibration with the well solutions was complete, and identifying drops containing a low precipitant concentration, that still remained clear. A drop containing crystals was also identified. After careful removal from their reservoirs, a cat's whisker was passed sequentially through the drop containing crystals and those that did not. It is thought that the microscopic aberrations on the surface of feline whiskers are especially adherent to micro-crystals, some of which are then dislodged when submerged in the subsequent pass through the unprecipitated crytsallisation drop. This was aimed to create a 'short cut' to the growth phase of the phase diagram shown in Figure 2.6 on page 26. By depositing small crystals into the pre-equilibrated drops directly, it removes the need for the supersaturation usually required for nucleation. This technique proved quite successful and increased the number of useable crystals by several fold. These were subsequently cryoprotected using lithium sulfate and glycerol as described previously and used in X-ray diffraction experiments with radiation of energy 13001.25 eV, on the PX2 protein crystallography beamline of the Australian Synchrotron, Melbourne, Australia. Further details of crytsallisation and data collection are summarised in Table 6.1 on the facing page. The resulting diffraction data was indexed, measured and scaled with the HKL2000 package [42]. Phase estimates were obtained via molecular replacement as implemented by the Phaser [128] utility in Phenix [45] using the previously published native structure (PDB identifier: 1EEM) [4] as a search model (Table 6.2 on page 142). Subsequent model building and refinement was performed alternatively with Coot [43] and the Phenix [45] programs, respectively.

6.2.1.2 Experimental conditions associated with the crytsallisation of hGSTO1 in an orthorhombic lattice.

The literature review regarding the crytsallisation of protein-ligand complexes by Hassell *et. al.,* [126] also suggested the use of a 'limited-additive screen', in which compounds known to increase ligand solubility and crystal quality are combined with previously known crytsallisation conditions. A screen of the suggested compounds using the concentrated

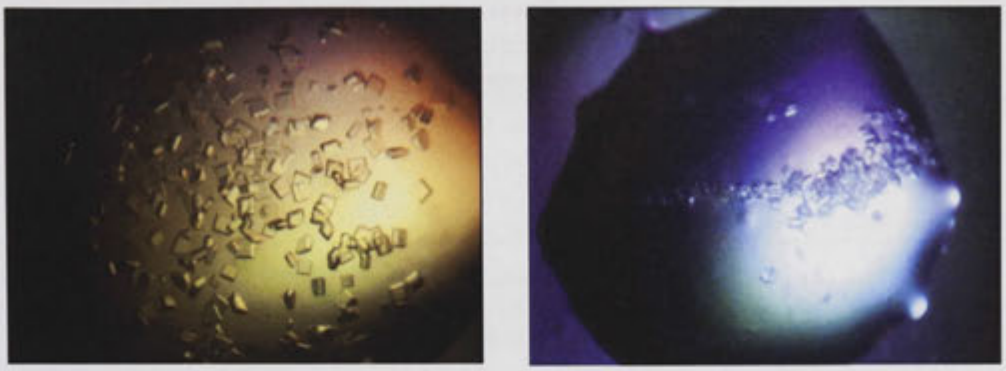


Figure 6.3
A redox sensitive purification protocol of hGSTO1 has resulted in the formation of the tetragonal crystals shown here, despite the crytsallisation conditions being analogous to those previously reported to induce crystal growth in a trigonal lattice [4]. Due to the relatively recalcitrant growth of diffraction quality crystals in this form (left) and the limited supply of protein, efforts were made to nucleate crystal growth in drops of low precipitant concentration via the method of 'streak seeding' (right), using a cat's whisker.

Crystallisation	Experimental parameters of hGSTO1-1 in complex with GSH in a tetragonal crystal lattice
Reservoir solution	1.7 M $(NH4)_2SO_4$
	100 mM Sodium Acetate pH 4.25
Crystallisation drop	1 µL hGSTO1-1 (25.18 mg/ml)$^{\mp}$
	1 µL reservoir
Cryoprotectant solution	2.2 M Li_2SO_4
	100 mM Sodium Acetate pH 4.75
	Sequential transfer from 5%——>10——>15%
	Glycerol solution
Data collection	
Resolution limits, Å	50.00-3.21 (3.32-3.21†)
Space group	$P4_222$
Unit Cell, Å	a=b=129.66 c=149.13
Total reflections	307 192
Unique reflections	21 224
Completeness, %	99.0 (93.5†)
I/σ	35.3 (6.0†)
Redundancy	14.5 (13.7†)
Refinement	
R_{cryst}/R_{free}, %	20.4/24.9
rmsd bonds, Å	0.011
rmsd angles, °	1.471
Number of waters	10
Ligands	GSH, glycerol, acetate, SO_4^{-2}, Cl^-

†, the highest resolution shell. $^{\mp}$, previously incubated with a molar excess of the CP-424174 inhibitor.

Table 6.1
Crystallisation details and structure determination statistics of the hGSTO1 native complex with GSH in space group $P4_222$.

Search Model	LLG (log-likelihood gain)	Rotation function Z-score	Translation function Z-score	Packing index
Protein monomer 1	293	6.1	22.9	0
Protein monomer 2	1059	7.5	40.1	0
Protein monomer 3	2118	5.4	40.9	0
Protein monomer 4	3179	4.9	37.5	0

Table 6.2
Molecular replacement statistics from Phaser [128] of the hGSTO1 native complex with GSH in space group $P4_222$.

batch of protein that had not been used in the dilute inhibitor incubation experiments (Section 6.2.1 on page 136) had resulted in a novel orthorhombic crystal form of hGSTO1-1 when the previously published $(NH_4)_2SO_4$/Na Acetate condition [4] was combined with 1 M guanidine hydrochloride. Like the tetragonal lattice discussed previously, this suggests it is the altered method of protein purification, and not the presence of inhibitors that facilitated crytsallisation in the new morphology. Application of analogous crytsallisation conditions to the 'dilute inhibitor incubation' protein preparations detailed above however (Section 6.2.1 on page 136), resulted in numerous crystals derived from all 3 of the inhibitor/protein solutions (Figure 6.4 on page 144). These were subsequently cryoprotected using a solution of lithium sulfate, sodium acetate and guanidine hydrochloride, with or without the addition of glycerol. X-ray diffraction experiments upon these crystals were also performed at the Australian synchrotron, using the same settings as above. The resulting datasets were also phased using molecular replacement (Table 6.4 on the next page) as implemented by Phaser [128] with PDB entry 1EEM as the search model. A structure derived from one of the best quality datasets, diffracting to a resolution of 1.8 Å, is presented below. Despite the excellent diffractive qualities of these crystals, subsequent analysis of the datasets with the phenix.xtriage utility of Phenix [45] revealed that all crystals of this space group were almost perfectly pseudo merohedrally twinned (Table 6.5 on page 144). Thus, Refmac [129] was used with amplitude based twinning refinement turned on in subsequent refinements, alternated with rounds of manual model building in Coot [43]. Further details about crytsallisation and data collection are summarised in Table 6.3 on the next page.

Crystallisation	Crystallisation and structure determination statistics of hGSTO1 native in a C-centered orthorhombic crystal lattice
Reservoir solution	2.1 M $(NH4)_2SO_4$ 100 mM Sodium Acetate pH 4.6 1 M Guanidine hydrochloride
Crystallisation drop	1 µL hGSTO1 (15.21 mg/ml)$^{\mp}$ 1 µL reservoir
Cryoprotectant solution	2.2 M Li_2SO_4 100 mM Sodium Acetate pH 4.75 1 M Guanidine hydrochloride
Data collection	
Resolution limits, Å	50.00-1.81 (1.87-1.81†)
Space group	$C222_1$
Unit Cell, Å	a=59.06 b=102.29 c=259.25
Total reflections	504 266
Unique reflections	95 970
Completeness, %	99.9 (99.2†)
I/σ	23.2 (1.3†)
Redundancy	5.3 (3.4†)
Refinement	
R_{cryst}/R_{free}, %	23.9/27.9
rmsd bonds, Å	0.012
rmsd angles, °	1.891
Number of waters	395
Ligands	GSH, glycerol, acetate, DMSO, Tris, SO_4^{-2}, $[CH_6N_3]^+$, Cl^-

†, the highest resolution shell. $^{\mp}$, previously incubated with a molar equivalent of CP-424174 (Figure 6.1b on page 134).

Table 6.3
Crystallisation details and structure determination statistics of the hGSTO1 native complex with GSH in space group $C222_1$.

Search Model	LLG	Rotation function Z-score	Translation function Z-score	Packing index
Protein monomer 1	232	13.5	12.9	0
Protein monomer 2	780	11.2	21.2	11

Table 6.4
Molecular replacement statistics from Phaser [128] of the hGSTO1 native enzyme in space group $C222_1$.

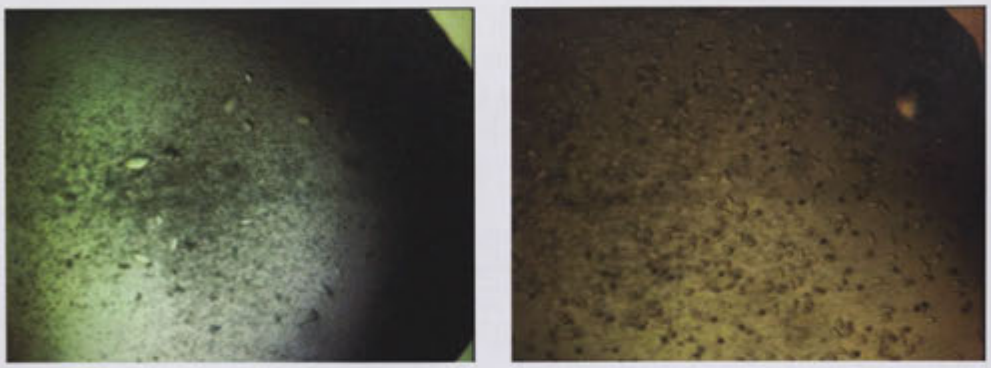


Figure 6.4
A limited additive screen [126] of the published conditions of hGSTO1 crytsallisation [4] combined with an altered method of purification has resulted in the new crystal morphology shown above. Despite excellent diffractive qualities, the nearly perfect pseudo merohedral twinning (Table 6.5) of this C-centered orthorhombic crystal lattice first needed to be overcome before the associated structure of hGSTO1 could be characterised (Table 6.3 on the previous page).

Twinning operator	Type	Estimated twin fraction (Britton analysis)	Estimated twin fraction (H-test)	Estimated twin fraction (Maximum likelihood method)
1/2*h-1/2*k,-3/2*h-1/2*k,-l	Pseudo merohedral	0.449	0.478	0.478
1/2*h+1/2*k,3/2*h-1/2*k,-l	Pseudo merohedral	0.449	0.478	0.478

Table 6.5
Twinning analysis statistics from phenix.xtriage [45].

6.3 Results and Discussion

The tetragonal and orthorhombic crystal lattices were discovered to contain 4 and 2 molecules within their asymmetric units respectively (Figure 6.5 on page 147). Although the tetragonal structure was solved at the modest resolution of 3.21 Å, the high level of NCS present was effectively employed to increase the observation to parameter ratio of the dataset, allowing the resolution of clearly defined electron density and good crystallographic R-factors (Table 6.1 on page 141). The four monomers were seen to be forming two pairs of physiological dimers (consisting of chains AC and BD) via 2 non-crystallographic two fold axes of symmetry. Conformational change was clearly observable to have destroyed the secondary structure of the C-terminal helix α9. The resulting random coil configuration within this region was unique to each, and the short helix α10 was completely disordered. Electron density within the 4 molecules terminated between residues 223-227 (Figure 6.6 on page 148). Since several residues in this region have previously been shown to play important roles in ligand binding within the H-site (Sections 5.3 on page 102 and 4.3 on page 71), this may reflect conformational plasticity that could have a specific purpose during catalysis. This hypothesis is corroborated by similar conformational changes we have observed in the hGSTO1 Δ155 isoform, that displays increased enzymatic activity *in vitro*, but appears to cause a deficiency *in vivo* due to an associated instability of the enzyme (Appendix A on page 181).

Within the active site of this tetragonal crystal form, GSH was observed forming a mixed disulfide with C32 as observed previously [4] within all four chains. While the mode of complex formation within the G-site was generally conserved, a slightly altered conformation of GSH was refined within the G-site of chain D, although this may be due to its conformation being insufficiently restrained by the poorly defined electron density (not shown) in this region relative to the other 3 chains (Figure 6.6 on page 148).

Crystallographic disorder of the C-terminal domain was also observed in the orthorhombic lattice, although in this instance, it appears to have been a prerequisite for crystal formation. Helices α9 and α10 of the hGSTO1 crystal structure previously published [4] overlap with one another when superimposed into this space group (Figure 6.7 on page 150), a physical restraint that is reflected in the high packing index reported by Phaser (Table 6.4 on page 143) that used this structure as the search model. The complete disorder of this region appears to be essential for crystal contacts of the asymmetric unit to occur. This crystal form was dependent on the presence of 1 M guanidine hydrochloride, a

potent denaturing agent of proteins. It is probable that this compound was able to induce crytsallisation in this space group due to its denaturing properties further disordering the already labile C-terminal helices, thereby allowing the necessary crystal contacts to occur. Although the inherent pseudo merohedral twinning of this crystal form has resulted in high crystallographic R-factors given the resolution (Table 6.3 on page 143), the quality of the electron density is unquestionable, allowing many structural features to be observed with certainty. Any inferences based on this crystal structure must be made with caution however, since the high concentration of a known protein denaturant may have artificially influenced its structural fold in ways that are not physiologically relevant. Nevertheless, its structural congruency with the structure previously published (rmsd of 0.41 over 207 residues) implies this is minimal within the regions resolved within electron density. While the active sites of both monomers were devoid of ligands, a guanidinium ion was observed tightly bound to the carboxylate of E85 in chain A (Figure 6.8 on page 151), while the C32 thiol of chain B appears to have been derivatized to the sulfenic acid derivative (Figure 6.9 on page 152), possibly by the X-ray beam.

6.3.1 Structural features of hGSTO1-1 within the tetragonal lattice, $P4_222$

Despite attempts to remove the tripeptide by dialysis prior to crytsallisation, all molecules within the asymmetric unit of the $P4_222$ crystal form were observed to be covalently bound to GSH, in an analogous confirmation to that observed within the previously published trigonal lattice [4]. Significant changes were observed within the adjacent H-site however, due to conformational heterogeneity within the C-terminus. This was observed in all 4 molecules of the asymmetric unit, but the structural changes observed were unique in each (Figure 6.6 on page 148). Despite completely different crystal contacts, the remainder of the structure is remarkably similar to the equivalent complex with GSH crystallised in the trigonal space group, $P3_121$ [4]. The observation of similar crystallographic disorder in this region within other space groups (Appendix A on page 181 and the orthorhombic lattice presented here) suggests this conformational flexibility may serve an important physiological role within the body.

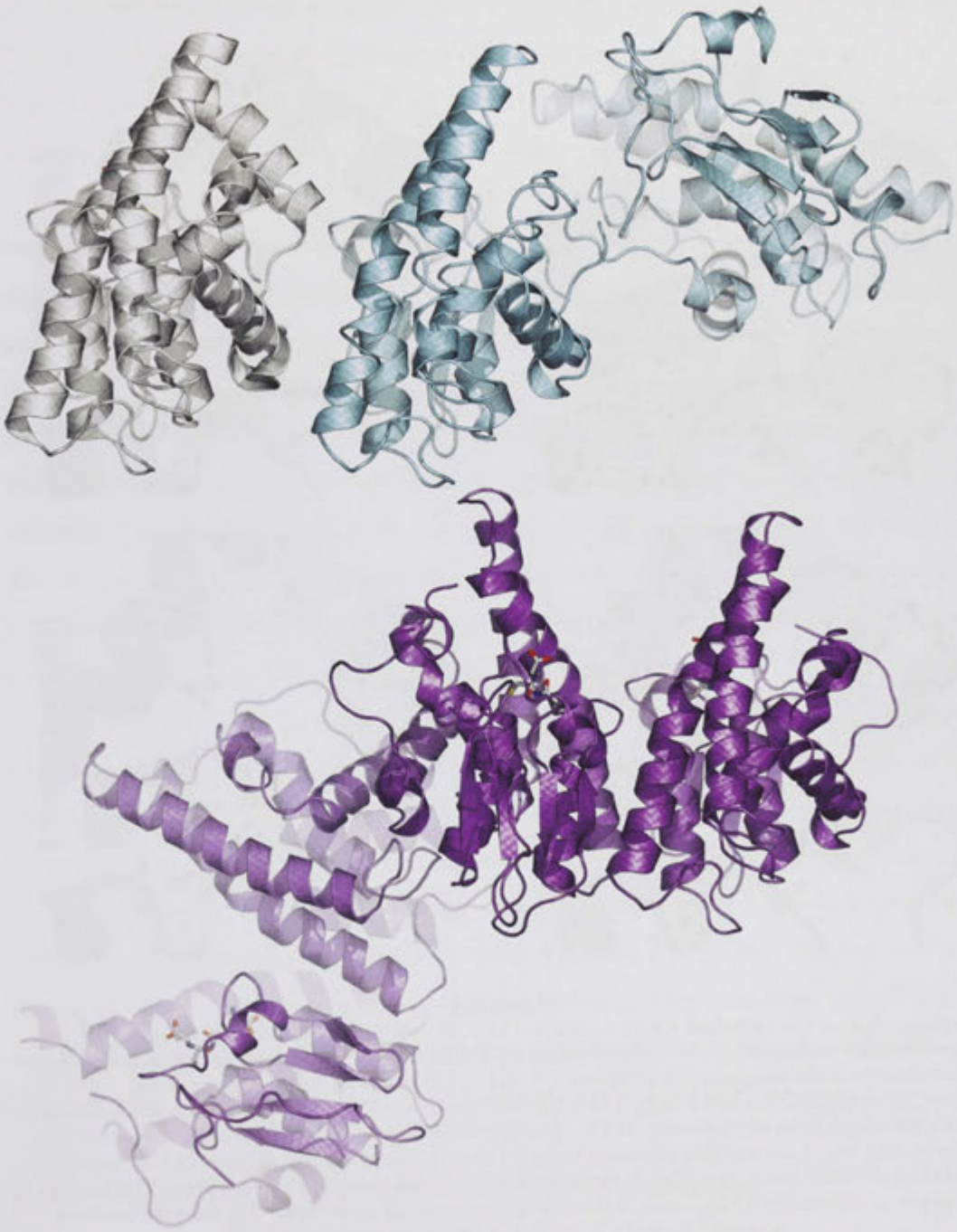


Figure 6.5
A comparison of the published structure of hGSTO1 (gray) [4] with the asymmetric units comprising the tetragonal (magenta) and C-centered orthorhombic (cyan) crystal structures presented here.

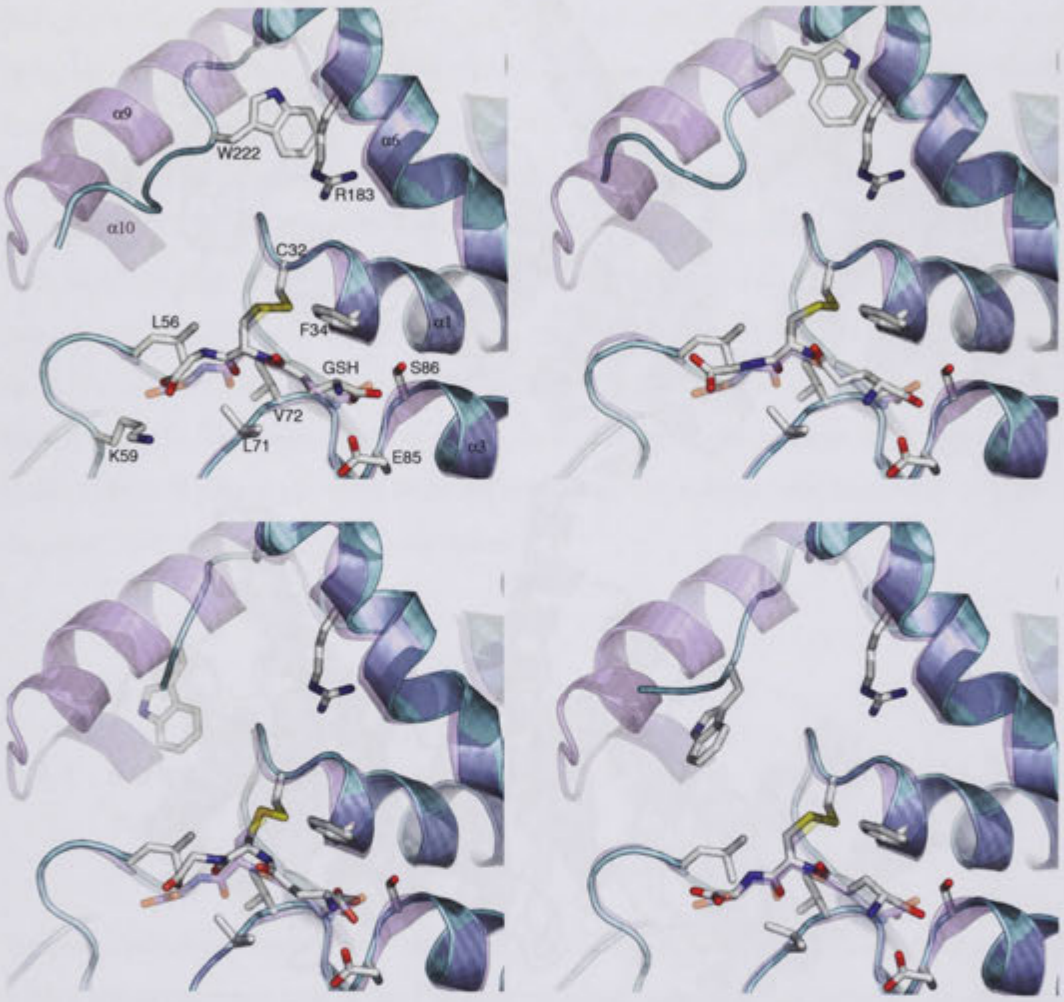


Figure 6.6

A comparison of the published $P3_121$ complex of hGSTO1 with GSH (translucent magenta) [4] with the equivalent regions of the 4 molecules within the $P4_222$ asymmetric unit (cyan). Shown clockwise from the top is the superposition of chains A (rsmd = 0.76 Å over 220 residues), B (rmsd = 0.83 Å over 217 residues), C (rmsd = 0.76 Å over 216 residues) and D (rmsd = 0.91 Å over 217 residues) with the coordinates of PDB entry 1EEM. The mixed disulfide observed with GSH, associated side chains, and two H-site residues identified to be involved in GSSG binding (Figure 4.11 on page 81), R182 and W222, are shown in stick representation, coloured according to atom type. The later is shown in translucent within chain D due it's side chain being disordered. NB: helix $\alpha 4$ has been omitted for clarity in viewing H-site conformational change.

6.3.2 Structural features of hGSTO1-1 within the C-centered orthorhombic lattice, $C222_1$

The high resolution crystal structure of hGSTO1 with a C-centered orthorhombic lattice displays an interesting crystal packing within its asymmetric unit that precludes the formation of helices α9-10 in its structure. The residues associated with these elements of secondary structure were not observed due to disorder. While the structural rearrangement was likely to have been initiated by the 1 M concentration of guanidinium hydrochloride present in the crytsallisation conditions, that fact that structural flexibility within this region has also been observed within other crystal forms of the enzyme (Appendix A on page 181 and the tetragonal crystal form discussed above) suggests it may have physiological relevance. In support of this hypothesis, the structural fold of the remaining polypeptide chains is remarkably similar to that observed in other crystal forms. The two chains within the $C222_1$ asymmetric unit, A and B, superimpose on the single molecule comprising the trigonal space group asymmetric unit [4] with rmsd values of 0.41 Å over 207 residues, and 0.47 Å over 204 residues, respectively. The novel crystal contacts of the $C222_1$ asymmetric unit, contrasted to the published structure of hGSTO1 is shown in more detail in Figure 6.7 on the following page.

It is possible that the apparent conformational change in the C-terminal extension may have affected the binding affinity of GSH and/or the inhibitor. Both G-sites within the asymmetric unit were found to be unexpectedly empty, despite these crystals coming from the same protein purification/dilute inhibitor incubation as that used to grow the tetragonal complex with GSH presented above. A guanidinium ion is observed tightly binding to the carboxylate of E85 within chain A (Figure 6.8 on page 151), while in chain B, the active site cysteine appears to have been derivatized to the sulfenic acid (Figure 6.9 on page 152). Given the care taken to ensure reducing conditions during purification, it is probable that this occurred subsequent to exposure within the X-ray beam.

6.3.3 Hypothesised mechanism of hGSTO1 inhibition

It was disappointing that despite every care being taken to remove potentially competitive ligands from the experimental conditions, and the subsequent development of a novel, redox sensitive purification protocol leading to the crytsallisation of hGSTO1 in two novel morphologies, the mode of hGSTO1 inhibition could not be confirmed crystallographically in the present study. Within lower resolution crystal structures of the $C222_1$ crystal form,

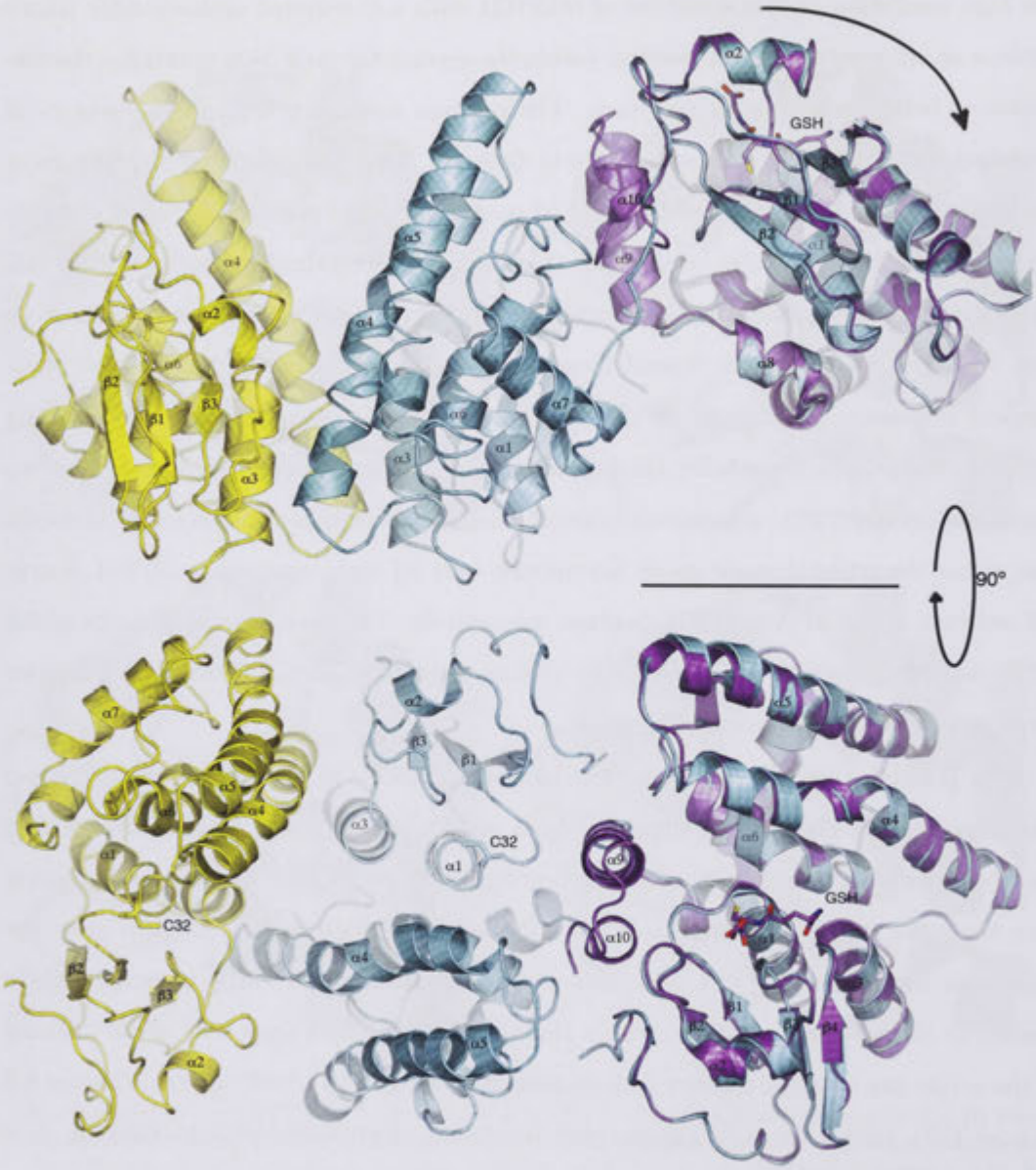


Figure 6.7

A representation of the novel mode of crystal packing within the recently discovered orthorhombic crystal form of hGSTO1. The asymmetric unit of space group $C222_1$ is shown in cyan with the previously published structure (PDB identifier: 1EEM), superimposed on chain B in magenta. When viewed from above, one can see that the two helices at the interface, $\alpha 9$ and $\alpha 10$, are incompatible with the current mode of crystal packing. In yellow is a crystallographically related symmetry partner that forms the physiological dimer with chain A of the asymmetric unit.

Figure 6.8

The observed active site residues of chain A within the $C222_1$ crystals of hGSTO1. The crystallographic fold of hGSTO1 within the orthorhombic lattice is shown as a cyan cartoon, overlaid in translucent magenta with the published trigonal lattice structure [4]. Both active sites were unexpectedly empty, suggesting associated conformational change in the C-terminus (helices $\alpha 9$ and $\alpha 10$) may have decreased this sites affinity for GSH (since co-purified protein crystallised in a GSH complex within the tetragonal lattice above). Active site residues are shown in stick representation coloured according to atom type, with associate $2mF_O - DF_C$ electron density depicted as gray isomesh, contoured at a level of 1.5 σ. A guanidinium ion was observed tightly bound to the carboxylate of E85.

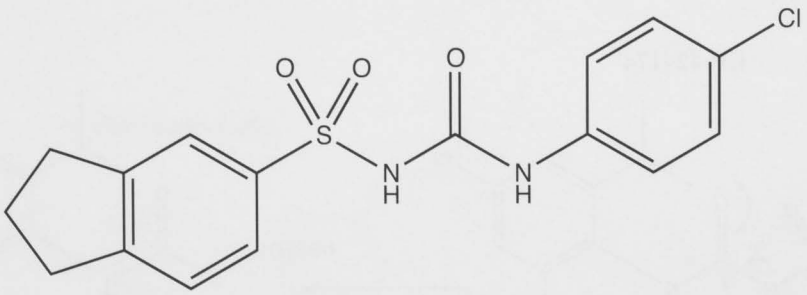
Figure 6.9
The observed active site residues of chain B within $C222_1$ crystals of hGSTO1, shown in a manner analogous to Figure 6.8 on the preceding page. The active site cysteine was observed to have been derivatized to the sulfenic acid, possibly via radiation damage in the X-ray beam.

residual 'blobs' of density were often observed in the G-site, indicating that inhibitors may have been bound at a low occupancy (not shown), however attempts to model these with logical fragments of the respective compounds were unsuccessful. In the 'dilute inhibitor incubation' experiments described previously (Section 6.2.1 on page 136), the final concentration of inhibitors (10 μM) was in a slight molar excess relative to the enzyme (7.6 μM). While this was deemed to be sufficient given the reportedly high affinity of these compounds for the hGSTO1 active site [32, 24], this technique may have proved more successful by significantly increasing this molar ratio. Based on the previous characterisation of enzymatic inhibition by similar compounds however, a logical hypothesis can be framed as to the mechanism by which hGSTO1 inhibition would proceed.

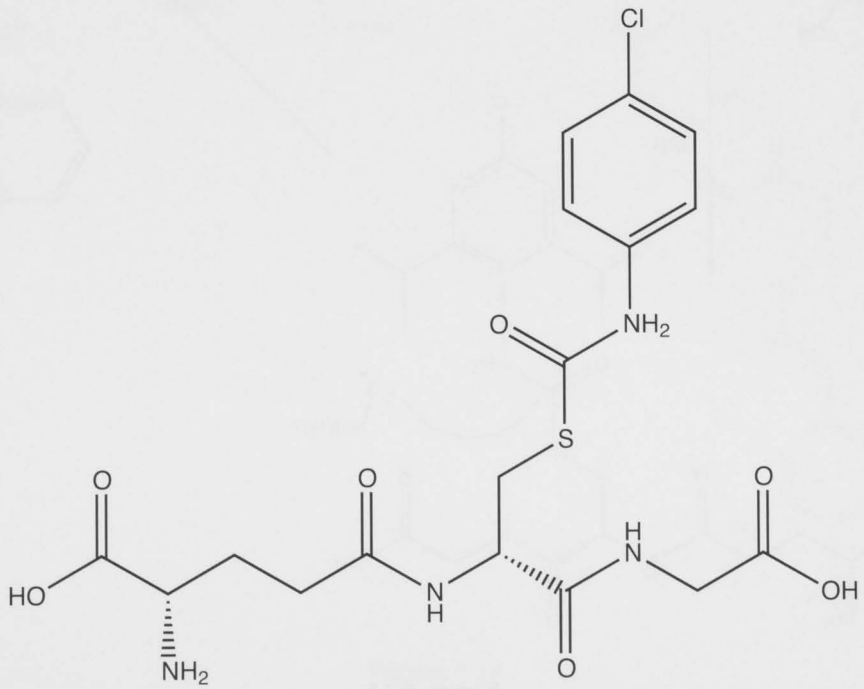
Perhaps the most insightful information about the activity of the diarylsulfonyl urea inhibitors developed by Pfizer (Figures 6.1a on page 134 and 6.2 on page 138) can be found in previous work that aimed to discover the mechanism of the anti-tumour agent Sulofenur [130], a clear progenitor of these next generation CRIDs [127]. These researchers showed that this compound was metabolised to the glutathione conjugate *in vivo*, and subsequently degraded to the mecapturic acid. Interestingly, they noted that the former was a potent inhibitor of glutathione reductase [130]. This glutathione conjugate (Figure 6.10 on page 155) was characterised by mass spectrometry [130], and could be expected to have an analogous structure to the corresponding active site cysteinyl, or free glutathionyl conjugates of the CRIDs CP-424174 and CP-457324 studied here. As noted previously, the glutathione conjugates of these compounds were also capable of inhibiting the post-translational processing of IL-1β [24]. Wether hGSTO1-1 is most effectively inhibited by the compounds themselves or their metabolical derivatives is unknown (Figure 6.11 on page 156).

The α-chloroacetamide inhibitor (Figure 6.1b on page 134) however, would be expected to exclusively inhibit hGSTO1 via a direct interaction with the enzyme (Figure 6.12 on page 157). It has been previously noted to function via covalent modification of the active site cysteine [32], and this would be expected to feature nucleophilic attack upon its chloroacetyle group. This is the most probable leaving group within its structure, and its small size would be incompatible with the precise selectivity with which it reportedly binds hGSTO1 if mediated thorough a glutathionyl conjugate intermediate. The polyaromatic functionalities would instead be expected to form tight hydrophobic interactions with several residues within the active site, such as F34, that would optimally orient its carbonyl

group for nucleophilic attack by C32. The resulting secondary amine would be protonated, leaving the catalytically essential thiol inactivated by covalent modification.



Sulofenur



Glutathione conjugate of Sulofenor

Figure 6.10
The anti-cancer compound Sulofenur and its glutathione conjugated metabolite [130].

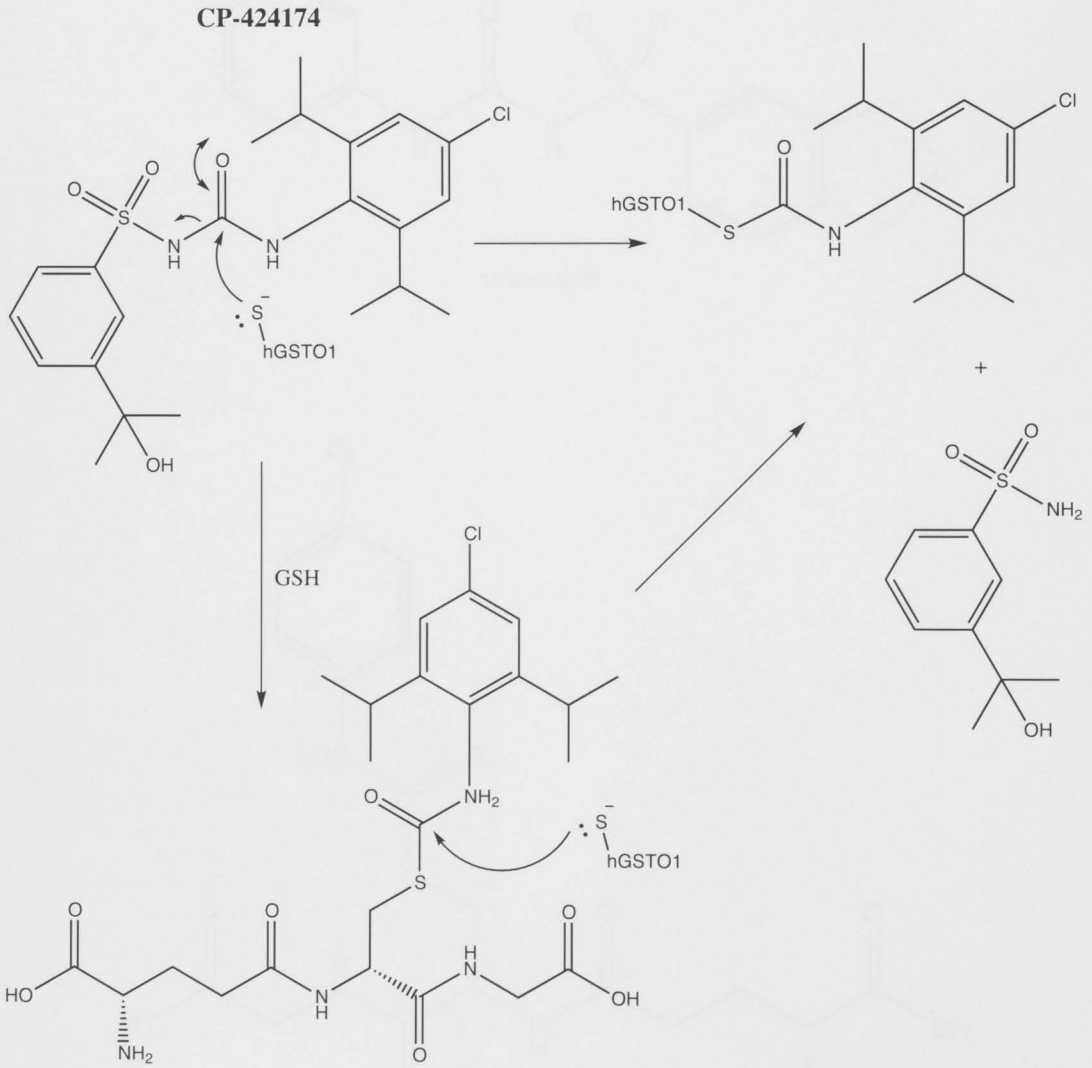


Figure 6.11
Hypothesised mechanisms of hGSTO1 inhibition by diarylsulfonyl ureas such as CP-424174. Covalent modification of the active site cysteine could occur via a direct interaction or a metabolically derived glutathione conjugate [24].

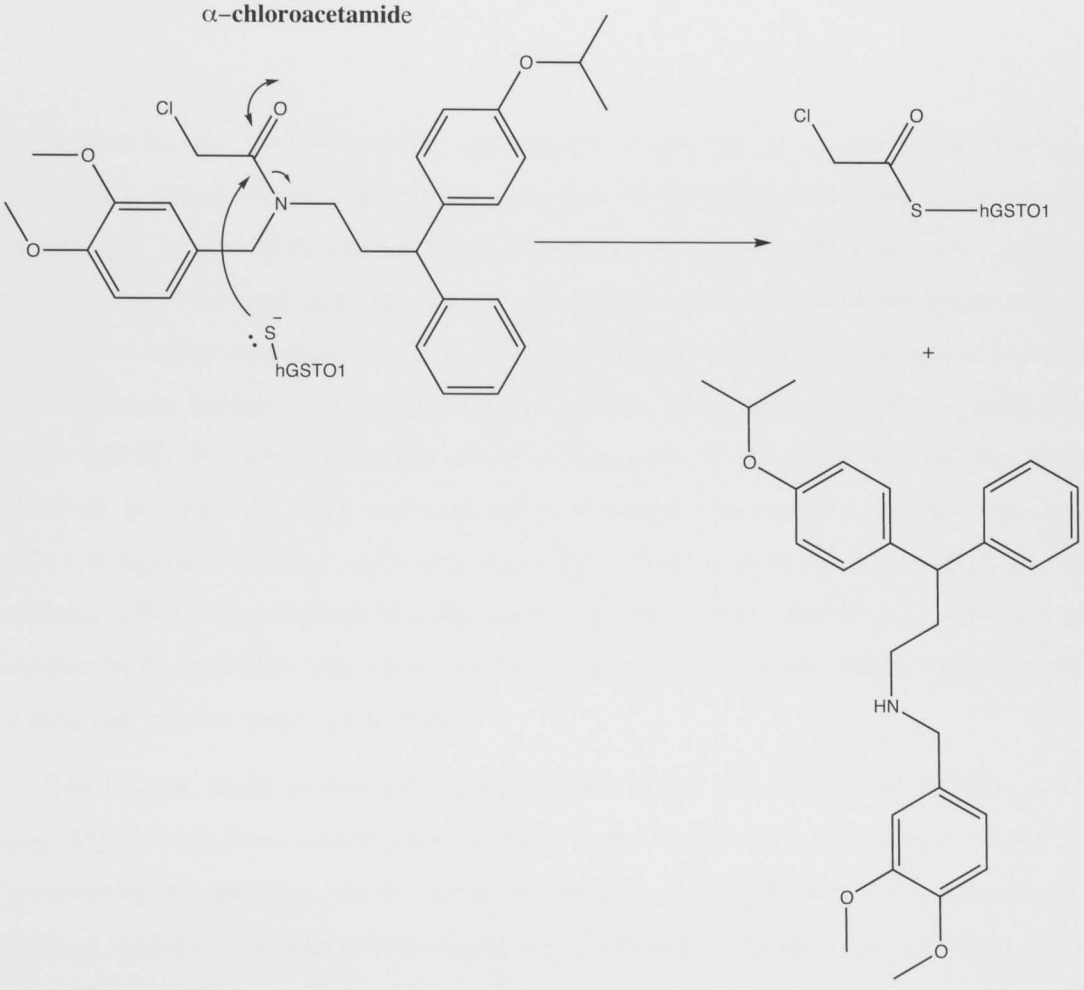


Figure 6.12
The inhibition of hGSTO1 by the α-chloroacetamide recently identified by Bachovchin *et. al.*, [32] is expected to proceed via the chloroacetylation of its active site cysteine residue.

Concluding remarks

In conclusion, the crystallographic investigation of mechanism in the human enzyme, hGSTO1-1, has yielded considerable insight into the relationships between its structural adaptability and the chemical reactions it performs. States associated with the catalysis of almost all of its small molecule substrates have been obtained in high resolution crystal structures. These include the binding of the product ascorbate within the active site and the concurrent binding of glutathionyl methylarsenite with a derivative of the model substrate CDNB. The ternary complex of the enzyme with GSSG and 4NPG has also been obtained, in which the latter was observed at the dimer interface in a position that may reflect a 'ligandin' binding mode or a preliminary docking position of the substrate. In addition, efforts to crystallographically characterise the mode of hGSTO1-1 inhibition has resulted in its structure being solved within two novel crystal forms. Each of these results is reflected upon in more detail below.

The binding mode of ascorbate within the active site was unexpected (Chapter 3 on page 43). Rather than revealing the position of the product at the conclusion of the dehydroascorbate reduction reaction cycle as intended, this binding mode is suggestive of product inhibition, in which intracellular ascorbate concentrations may self regulate *in vivo*. This conclusion is evidenced by the mutually exclusive mode of ascorbate binding with the formation of a cysteine-GSH mixed disulfide, and the lack of planarity and double bond conjugation in the superposed substrate [48], both of which would make the observed π-stacking interaction with F34 unfavourable. It's isozyme hGSTO2-2, has a much higher activity towards this substrate and is in fact, the most efficient mammalian dehydroascorbate reductase characterised to date [25]. This implies that the findings within this research may have important physiological applications if such a binding mode was found to be conserved between the two. Indeed, a comparison of the region associated with ascorbate binding (Figure 3.10 on page 59) suggests this to be the case. This superposition identifies

H71 as the only non-conservative amino acid substitution within the equivalent region of hGSTO2. Its replacement of a leucine in the active site of hGSTO1 may convey an associated decrease in affinity of this inhibitory binding mode, contributing to the higher level of enzymatic activity observed. Conversely, its capacity to act as a general base within the reaction mechanism may also contribute to the increase in catalytic efficiency observed. This reaction has typically been thought to proceed via a ping-pong mechanism involving the formation of a covalent cysteinyl-thioketal intermediate, that collapses to yield the reduced product and a mixed disulfide upon nucleophilic attack by GSH (Figure 3.1 on page 44). This is the mechanism proposed by researchers studying this reaction as catalysed by the DHAR of spinach chloroplasts, who also found evidence of product inhibition by ascorbate during their kinetic analysis of the enzyme [54]. Substantial evidence has been presented for a different mechanism however, based on the characterization of a DHAR enzyme of *A. thaliana* [56] that shows significant sequence similarity to hGSTO1 (Figure 3.8 on page 56). These authors could not observe the hypothesised cysteinyl-thioketal intermediate by ESI-TOF MS, and pointing out that in chemical systems analogous to dehydroascorbate, the attack of a cysteine thiol would lead to a dead-end thioketal-enzyme complex, suggest a SET mechanism instead (Figure 3.7 on page 55). Such a mechanism would be consistent with the well established ability of thiol compounds to act as hydride donors during radical reactions [56]. Further research characterising this reaction within the Omega class should involve a thorough kinetic analysis that investigates both the mode of product inhibition inferred by the structural findings presented here, and evaluates the actuality of either a ping-pong or SET mechanism of dehydroascorbate reduction. The latter may also be investigated by carefully controlled electron paramagnetic resonance (EPR) experiments with dehydroascorbate in the presence and absence of the enzyme. Site directed mutagenesis experiments of H71 within hGSTO2-2 are also warranted in order to quantitate its contribution to the higher activity of this enzyme relative to hGSTO1-1. The altered catalytic turnover of this substrate associated with genetic polymorphisms may be the causative factor implicating the omega class in the AAO of Alzheimer's and Parkinson's disease in genetic mapping studies [22, 49]. Both of these conditions have been associated with oxidative stress, to which ascorbate plays a key protective role within the body and in particular, the brain, where its concentration is most tightly regulated [50].

Perhaps the most interesting finding of this investigation into the mechanism of hGSTO1 was the identification of a glutathionylated methyl arsenical bound within its

active site (Figure 4.6 on page 75). This supports previous findings that glutathionylated, trivalent methyl arsenicals are the required substrate for the hAS3MT enzyme [71], and helps resolve conjecture in the literature regarding the identity of intermediates within this metabolic pathway [25, 68, 69, 70]. This information is crucial to obtaining a more detailed understanding of the pathomechanisms of chronic arsenic toxicity currently affecting approximately 100 million people in the geographical region surrounding Bangladesh [75], and to obtaining a more fundamental understanding of the pharmacology of ATO in its treatment of APL in combination with ATRA [78]. While initial investigations indicated that the cytotoxic affect of this compound may be mediated via the generation of ROS [81, 82], it has since been conclusively shown to affect apoptosis via a direct interaction with the 'zinc finger' region within the PML domain. Together with ATRA binding to the RARα half of the fusion protein, it is thought that there is an associated conformational change that favours SUMOylation, degradation, and subsequent abolition of its inhibitory effect on cellular differentiation (Figure 4.2 on page 64) [80]. Since it is possible that ATO may affect these changes indirectly, via metabolites such as the glutathionyl methylarsenite species observed here, their identification may prove crucial to obtaining a better understanding of this interaction. Such information may also aid in the application of this treatment to other forms of cancer. PML has been implicated to play a vital role in maintaining quiescent populations of LICs implicated as the source of many relapses [85]. It has also been shown to be a key regulator of the p53 tumour suppressor, mutations in which are the most common genetic event in the formation of metastatic human cancers. PML has been found to actively upregulate the transcription of mutant p53 and is an important factor in the oncogenic 'gain of function properties exhibited by this protein [86]. In addition, a glutathione derivative of trivalent arsenic, *p*-4-[N -(S-glutathionylacetyl)amino]phenylarsenoxide, is currently undergoing clinical trials as a potent anti-cancer agent [87]. Its efficacy may now be reconciled in terms of it being a structural analogue of the metabolic intermediate that has currently been observed crystallographically. A thorough understanding of the catalytic mechanism associated with methyl arsonate reduction could benefit from the unequivocal determination of the binding order of substrates. This could be achieved by careful kinetic analysis of the reaction, however given the close proximity of two tryptophan residues in the active site (W180 and W222), fluorescence experiments may also aid the elucidation of this reactions binding sequence. Indeed, a similar technique has recently been applied to gain a more funda-

mental understanding of the mechanism of an arsenate reductase from *E. coli* [101]. In addition, the hydroxyl group of Y229 was observed to be participating in a solvent mediated hydrogen bonding interaction with the lone pair of arsenic (Figure 4.6 on page 75). This residue is located within the labile C-terminal helix α9, and may be the best candidate for the as yet unidentified hydrogen bond donor of the Omega class, observed within other GST classes to stabilise the GSH associated thiolate during catalysis. This tightly conserved residue is therefore also a prime candidate for future site directed mutagenesis studies. While an associated Y229F mutagenic PCR reaction was performed during this research, the time constraints of this project precluded the resulting DNA product from being sequenced, and the associated isozyme being characterised.

Serendipitously, the structure revealing bound glutathionyl methylarsenite also revealed the binding of HDNB, a derivative of the historical GST substrate CDNB. While subsequently confirmed by the binding of GSSG, this was the first confirmation of the predicted hGSTO1-1 H-site, and its binding was observed to be associated with several aspects of structural change. Alternate conformers of R132 and W222 were observed, the former of which was seen to be stabilised by a novel salt bridge with a sulfate ion (Figure 4.8 on page 77). It also appears to have triggered a 'conformational switch' in the adjacent α6-7 loop (Figure 4.9 on page 78), that was also corroborated by the glycinyl moiety of GSSG binding in the same position (Figure 5.5 on page 108). This has provided valuable insight into the structural dynamics of the hGSTO1 H-site, and may represent a structural adaptation that could be essential for efficient catalytic turnover and enzymatic promiscuity.

This latter binding mode of GSSG was observed in a ternary complex of hGSTO1 and 4NPG. In lieu of the phenacyl glutathione substrate, this disulfide was observed in the active site instead, and is the first time its binding mode has been confirmed crystallographically, despite it being a hypothesised product of all hGSTO1-1 mediated chemistry. Comparison with other complexes of GSTs with GSSG in the PDB (Figures 5.10 on page 115 and 5.11 on page 116) revealed a tightly conserved binding mode of the G-site bound half of the molecule, that corresponded well with the previously observed binding mode of GSH [4]. The conformational space sampled by the other half of the molecule however, was seen to be of a much greater variety. The evolution of a relatively wide dimer interface within the Omega class appears to have precluded inter-dimeric interactions with the product molecule while the appendage of its two C-terminal helices (α9-10) has resulted in a much deeper H-site. This is observed to have created a much 'tighter' binding mode

within hGSTO1 relative to the other two examples. In the binding mode observed here, the H-site bound half of the molecule is actually orientated away from, rather than towards, the interdimeric cleft. This is expected to have profound implications for this products binding affinity, and subsequently, for all aspects of hGSTO1-1 mediated chemistry.

Conversely, 4NPG was found to bind along the crystallographic two fold corresponding to the hGSTO1-1 dimer interface. While repeated attempts failed to observe its binding mode within the active site, the binding of HDNB suggested its nitrophenacyl moiety could be accommodated in a similar position, as did the molecular modeling of the compound into the electron density associated with bound GSSG (Figure 5.15 on page 122). This crystallographically observed position may be representative of a 'ligandin' binding mode. Indeed, similar examples of ligand binding at the dimer interface have been observed crystallographically in other GSTs (Figure 5.9 on page 113). The mutually inclusive nature of 4NPG's binding with GSSG, coupled with their close proximity within the crystal structure however, suggests this may be a preliminary docking position of the substrate. Repeated attempts to observe both of these compounds independently have been unsuccessful, suggesting that this crystal structure may feature as part of the physiological reaction cycle, with 4NPG moving concertedly into the active site with GSSG departure.

Finally, efforts to discover the mode of inhibition by diarylsulfonyl urea CRIDs [24] and an α-chloroacetamide [32] have ultimately proved unsuccessful but have allowed the structure of hGSTO1-1 to be solved in two new crystal forms. Strikingly, the first of these, that is comprised of the tetragonal space group $P4_222$, crystallised from conditions that were directly analogous to those previously published to yield crystals of a trigonal lattice [4]. The only variable being the conditions associated with purification of the protein (Section 6.2.1 on page 136), it would appear that the altered redox conditions associated with this new method were sufficient to induce a change in crystal morphology. The resulting structure was subsequently found to contain 4 monomers, comprising 2 physiological dimers, within its asymmetric unit. While the best diffraction obtained was to a modest resolution of 3.2 Å, the high level of NCS was effectively utilised to obtain clear electron density in which to model the polypeptide chains. A structural change was observed in the C-terminal helix α9 that was unique to each monomer, however the residues associated with the short α10 helix previously observed to follow [4] was completely disordered in all of them (Figure 6.6 on page 148). This crystal structure appears to represent a structurally dynamic state of the enzyme that lies at a position upon the potential energy landscape that is distinct from the

state previously crystallised in space group $P3_121$ [4]. This state apparently has a much higher affinity for GSH, since every molecule of the asymmetric unit was observed forming a mixed disulfide with the tripeptide despite attempts to remove it from the experimental conditions by dialysis prior to crystallisation (Section 6.2.1 on page 136). Furthermore, it would appear that the energy level represented by this crystal structure does not have a predisposition to a C-terminal extension with ordered secondary structure.

The second novel crystal form, discovered to belong to a C-centered orthorhombic lattice of space group $C222_1$, possessed much better diffractive qualities, but was unfortunately found to be pseudo merohedrally twinned. After overcoming this problem via amplitude based twinning refinement within Refmac [44], the resulting structure was also found to feature a dramatic conformational change within its C-terminus. This time it actually appeared to be a pre-requisite for crystal formation however, with the observed crystal contacts completely precluding the formation of helices $\alpha 9$ and $\alpha 10$. While the presence of 1 M guanidine hydrochloride was necessary for crystallisation in this morphology and undoubtedly contributed to the C-terminal crystallographic disorder observed, it is likely this structural change has a physiological basis. Taken together with highly similar structural plasticity observed in the tetragonal space group presented here and the structure of the hGSTO1Δ155 isoform we have recently published (Appendix A on page 181), there is compelling evidence that C-terminal conformational change is an essential physiological characteristic of hGSTO1-1 mediated catalysis. Since several residues in this region have been identified in the current study to be essential for ligand binding within the H-site, it is probable that this conformational change is associated with efficient entry and/or release of these compounds from the active site.

It is possible that the variety of conformational space sampled by the C-terminus of hGSTO1 within its various crystal forms is representative of different enzymatic conformations that are separated by small differences in free energy. The variations in the solvent environments associated with each crystal form is selective for each. These neighbouring states within the free energy landscape may feature as part of an 'induced fit' mechanism, by which the binding of certain substrates favour particular conformations. Indeed, this phenomena has been observed previously in the human GST α class, in which a similar, novel C-terminal helix is dependent on ligand binding in order to assume an ordered secondary structure [131].

While the size of the hGSTO1-1 physiological dimer (55 kDa) has historically been

considered too large for analysis by NMR spectroscopy, recent advances in both the resolving power of spectrometers and experimental techniques may allow this conformational change to be quantitated in concert with reaction of substrates in solution. The technique of Lanthanide tag labeling of proteins using the cell free synthesis incorporation of unnatural amino acids currently being developed by my colleagues [132, 133] is one such technique that may allow the structural plasticity of this region to be characterized within the context of enzymatic catalysis.

It is disappointing that despite the development of a new purification protocol that has lead to the discovery of two new crystal forms, the mode of hGSTO1-1 inhibition has been unable to be characterised crystallographically. With more time devoted to this research project, it is likely that this could have been achieved. Further research aimed at uncovering the mechanism by which these compounds inhibit hGSTO1-1 would do well to repeat the 'dilute inhibitor incubation' experiments outlined in Section 6.2.1 on page 136, but with an increased concentration of inhibitors and varied levels of GSH. The purification of the C32A and C32S active site mutations using this new purification protocol and subsequent analysis within these novel crystal forms may also prove invaluable in gaining further insight into the chemical mechanisms employed by this enzyme.

Bibliography

[1] Liska DJ. The detoxification enzyme systems. Altern Med Rev. 1998 Jun;3(3):187–98.

[2] Jeffery EH. Detoxification basics. Altern Ther Health Med. 2007 Jan;13(2):S96–7.

[3] Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. Biochem J. 2001 Nov;360(Pt 1):1–16.

[4] Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiin LS, Schulte GK, et al. Identification, characterization, and crystal structure of the Omega class glutathione transferases. J Biol Chem. 2000 Aug;275(32):24798–806.

[5] Gu Y, Guo J, Pal A, Pan SS, Zimniak P, Singh SV, et al. Crystal structure of human glutathione S-transferase A3-3 and mechanistic implications for its high steroid isomerase activity. Biochemistry. 2004;43(50):15673–15679.

[6] Patskovsky Y, Patskovska L, Almo SC, Listowsky I. Transition state model and mechanism of nucleophilic aromatic substitution reactions catalyzed by human glutathione S-transferase M1a-1a. Biochemistry. 2006;45(12):3852–3862.

[7] Oakley AJ, Bello ML, Battistoni A, Ricci G, Rossjohn J, Villar HO, et al. The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution1. Journal of molecular biology. 1997;274(1):84–100.

[8] Rossjohn J, McKinstry WJ, Oakley AJ, Verger D, Flanagan J, Chelvanayagam G, et al. Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site. Structure. 1998;6(3):309–322.

[9] Kanaoka Y, Ago H, Inagaki E, Nanayama T, Miyano M, Kikuno R, et al. Cloning and crystal structure of hematopoietic prostaglandin D synthase. Cell. 1997;90(6):1085–1095.

[10] Polekhina G, Philip G, Blackburn AC, Parker MW. Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity. Biochemistry. 2001;40(6):1567–1576.

[11] Armstrong RN. Structure, catalytic mechanism, and evolution of the glutathione transferases. Chem Res Toxicol. 1997 Jan;10(1):2–18.

[12] Habig W, Pabst MJ, Jakoby W. Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. The Journal of biological chemistry. 1974;249(22):7130.

[13] Ricci G, Bello ML, Caccuri AM, Pastore A, Nuccetelli M, Parker MW, et al. Site-directed mutagenesis of human glutathione transferase P1-1. Journal of Biological Chemistry. 1995;270(3):1243.

[14] Whitbread A, Tetlow N, Eyre H, Sutherland G, Board P. Characterization of the human Omega class glutathione transferase genes and associated polymorphisms. Pharmacogenetics and genomics. 2003;13(3):131.

[15] Board P, Anders M. Glutathione transferase omega 1 catalyzes the reduction of S-(phenacyl) glutathiones to acetophenones. Chem Res Toxicol. 2007;20(1):149–154.

[16] Mukherjee B, Salavaggione O, Pelleymounter L, Moon I, Eckloff B, Schaid D, et al. Glutathione S-transferase omega 1 and omega 2 pharmacogenomics. Drug Metab Dispos. 2006 Jul;34(7):1237–46.

[17] Whitbread AK, Masoumi A, Tetlow N, Schmuck E, Coggan M, Board PG. Characterization of the omega class of glutathione transferases. Methods in Enzymology. 2005 Jan;401:78–99.

[18] Paiva L, Marcos R, Creus A, Coggan M, Oakley A, Board P. Polymorphism of glutathione transferase Omega 1 in a population exposed to a high environmental arsenic burden. Pharmacogenetics and genomics. 2008;18(1):1.

[19] Tanaka-Kagawa T, Jinno H, Hasegawa T, Makino Y, Seko Y, Hanioka N, et al. Functional characterization of two variant human GSTO 1-1s (Ala140Asp and Thr217Asn). Biochem Biophys Res Commun. 2003 Feb;301(2):516–20.

[20] Marahatta S, Punyarit P, Bhudisawasdi V, Paupairoj A, Wongkham S, Petmitr S. Polymorphism of glutathione S-transferase omega gene and risk of cancer. Cancer Lett. 2006 May;236(2):276–81.

[21] Kolsch H, Linnebank M, Lutjohann D, Jessen F, Wullner U, Harbrecht U, et al. Polymorphisms in glutathione S-transferase omega-1 and AD, vascular dementia, and stroke. Neurology. 2004 Dec;63(12):2255–60.

[22] Li Y, Scott W, Zhang L, Lin P, Oliveira S, Skelly T, et al. Revealing the role of glutathione S-transferase omega in age-at-onset of Alzheimer and Parkinson diseases. Neurobiol Aging. 2006 Aug;27(8):1087–93.

[23] Kolsch H, Larionov S, Dedeck O, Orantes M, Birkenmeier G, Griffin W, et al. Association of the Glutathione S-transferase Omega-1 Ala140Asp Polymorphism With Cerebrovascular Atherosclerosis and Plaque-Associated Interleukin-1 alpha Expression. Stroke. 2007;38(10):2847.

[24] Laliberte R, Perregaux D, Hoth L, Rosner P, Jordan C, Peese K, et al. Glutathione S-transferase omega 1-1 is a target of cytokine release inhibitory drugs and may be responsible for their effect on interleukin-1β posttranslational processing. Journal of Biological Chemistry. 2003;278(19):16567.

[25] Schmuck E, Board P, Whitbread A, Tetlow N, Cavanaugh J, Blackburn A, et al. Characterization of the monomethylarsonate reductase and dehydroascorbate reductase activities of Omega class glutathione transferase variants: implications for arsenic metabolism and the age-at-onset of Alzheimer's and Parkinson's diseases. Pharmacogenetics and genomics. 2005;15(7):493.

[26] Yu L, Kalla K, Guthrie E, Vidrine A, Klimecki W. Genetic variation in genes associated with arsenic metabolism: glutathione S-transferase omega 1-1 and purine nucleoside phosphorylase polymorphisms in European and indigenous Americans. Environ Health Perspect. 2003 Aug;111(11):1421–7.

[27] Schmuck E, Cappello J, Coggan M, Brew J, Cavanaugh JA, Blackburn AC, et al. Deletion of Glu155 causes a deficiency of glutathione transferase Omega 1-1 but does not alter sensitivity to arsenic trioxide and other cytotoxic drugs. The International Journal of Biochemistry & Cell Biology. 2008;40(11):2553–2559.

[28] Dulhunty A, Gage P, Curtis S, Chelvanayagam G, Board P. The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator. Journal of Biological Chemistry. 2001;276(5):3319.

[29] Wang L, Xu J, Ji C, Gu S, Lv Y, Li S. Cloning, expression and characterization of human glutathione S-transferase Omega 2. International Journal of Molecular Medicine. 2005 Jan;16(1):19–27.

[30] Lillig CH, Berndt C, Holmgren A. Glutaredoxin systems. Biochim Biophys Acta. 2008 Nov;1780(11):1304–17.

[31] Kortemme T, Creighton TE. Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pKa values in proteins of the thioredoxin family. J Mol Biol. 1995 Nov;253(5):799–812.

[32] Bachovchin D, Brown S, Rosen H, Cravatt B. Identification of selective inhibitors of uncharacterized enzymes by high-throughput screening with fluorescent activity-based probes. Nature biotechnology. 2009;27(4):387–394.

[33] Ji X, Armstrong RN, Gilliland GL. Snapshots along the reaction coordinate of an SNAr reaction catalyzed by glutathione transferase. Biochemistry. 1993 Dec;32(48):12949–54.

[34] Cromer B, Morton C, Board P, Parker M. From glutathione transferase to pore in a CLIC. European Biophysics Journal. 2002;31(5):356–364.

[35] Adam GC, Sorensen EJ, Cravatt BF. Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. Nature biotechnology. 2002;20(8):805–809.

[36] Yan X, Pan L, Yuan Y, Lang J, Mao N. Identification of platinum-resistance associated proteins through proteomic analysis of human ovarian cancer cells and their platinum-resistant sublines. Journal of proteome research. 2007;6(2):772–780.

[37] Catanzariti AM, Soboleva TA, Jans DA, Board PG, Baker RT. An efficient system for high-level expression and easy purification of authentic recombinant proteins. Protein Sci. 2004 May;13(5):1331–1339.

[38] Astatke M. Comparison of HisDetectorTM Nickel-NTA Conjugates with a Single-Step Antibody Method for the Detection of His-Tagged Proteins. Nature Methods | Application Notes. 2008 May;.

[39] Rhodes G. Crystallography made crystal clear : a guide for users of macromolecular models / Gale Rhodes;. C2006 QP519.9.X72 R48 2006.

[40] Matthews B. The determination of the position of anomalously scattering heavy atom groups in protein crystals. Acta Crystallographica. 1966;20(2):230–239.

[41] Guss JM, Merritt EA, Phizackerley RP, Hedman B, Murata M, Hodgson KO, et al. Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic" blue" copper protein from cucumbers. Science. 1988;241(4867):806.

[42] Otwinowski Z, Minor W. [20] Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. 1997;276:307–326.

[43] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004 Dec;60(Pt 12 Pt 1):2126–32.

[44] Murshudov G, Vagin A, Dodson E. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997 May;53(Pt 3):240–55.

[45] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography. 2010;66(2):213–221.

[46] Davis I, Leaver-Fay A, Chen V, Block J, Kapral G, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W375–83.

[47] Washburn MP, Wells WW. The catalytic mechanism of the glutathione-dependent dehydroascorbate reductase activity of thioltransferase (glutaredoxin). Biochemistry. 1999;38(1):268–274.

[48] Kerber RC. " As Simple as Possible, but Not Simpler" The Case of Dehydroascorbic Acid. Journal of Chemical Education. 2008;85(9):1237.

[49] Li Y, Oliveira S, Xu P, Martin E, Stenger J, Scherzer C, et al. Glutathione S-transferase omega-1 modifies age-at-onset of Alzheimer disease and Parkinson disease. Human molecular genetics. 2003;12(24):3259.

[50] Rice M. Ascorbate regulation and its neuroprotective role in the brain. Trends in neurosciences. 2000;23(5):209–216.

[51] Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research. 2003;31(13):3784.

[52] Sampayo-Reyes A, Zakharyan R. Tocopherol esters inhibit human glutathione S-transferase omega. Acta Biochim Pol. 2006;53(3):547–52.

[53] Sampayo-Reyes A, Zakharyan R. Inhibition of human glutathione S-transferase omega by tocopherol succinate. Biomed Pharmacother. 2006 Jun;60(5):238–44.

[54] Shimaoka T, Miyake C, Yokota A. Mechanism of the reaction catalyzed by dehydroascorbate reductase from spinach chloroplasts. European Journal of Biochemistry. 2003;270(5):921–928.

[55] Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research. 2011;39(suppl 1):D225.

[56] Dixon DP, Davis BG, Edwards R. Functional divergence in the glutathione transferase superfamily in plants. Journal of Biological Chemistry. 2002;277(34):30859.

[57] Doyle M, Griffin J, Chinn M. Rearrangements of Ylides Generated from Reactions of Diazo Compounds Acceleration of the [2,3]-Sigmatropic Rearrangementwith Allyl Acetals and Thioketals by Catalytic Methods. Heteroatom. The Journal of Organic Chemistry. 1984 Jan;49:1917–1924.

[58] Jung CH, Wells WW. Spontaneous Conversion of L-Dehydroascorbic Acid to L-Ascorbic Acid and L-Erythroascorbic Acid. Archives of biochemistry and biophysics. 1998;355(1):9–14.

[59] Shimaoka T, Yokota A, Miyake C. Purification and characterization of chloroplast dehydroascorbate reductase from spinach leaves. Plant and Cell Physiology. 2000;41(10):1110.

[60] Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947.

[61] Challenger F. Biological Methylation. Chemical reviews. 1945;36(3):315–361.

[62] Crecelius E. Changes in the chemical speciation of arsenic following ingestion by man. Environmental Health Perspectives. 1977;19:147.

[63] Aposhian H, Gurzau E, Le X, Gurzau A, Healy S, Lu X, et al. Occurrence of monomethylarsonous acid in urine of humans exposed to inorganic arsenic. Chem Res Toxicol. 2000;13(8):693–697.

[64] Mandal B, Ogra Y, Suzuki K. Identification of dimethylarsinous and monomethylarsonous acids in human urine of the arsenic-affected areas in West Bengal, India. Chem Res Toxicol. 2001;14(4):371–378.

[65] Francesconi K, Tanggaar R, McKenzie C, Goessler W. Arsenic metabolites in human urine after ingestion of an arsenosugar. Clinical chemistry. 2002;48(1):92.

[66] Radabaugh TR, Sampayo-Reyes A, Zakharyan RA, Aposhian HV. Arsenate reductase II. Purine nucleoside phosphorylase in the presence of dihydrolipoic acid is a route for reduction of arsenate to arsenite in mammalian systems. Chemical research in toxicology. 2002;15(5):692–698.

[67] Guerin T, Astruc A, Astruc M. Speciation of arsenic and selenium compounds by HPLC hyphenated to specific detectors: a review of the main separation techniques. Talanta. 1999;50(1):1–24.

[68] Hansen H, Raab A, Jaspars M, Milne B, Feldmann J. Sulfur-containing arsenical mistaken for dimethylarsinous acid [DMA (III)] and identified as a natural metabolite in urine: major implications for studies on arsenic metabolism and toxicity. Chem Res Toxicol. 2004;17(8):1086–1091.

[69] Aposhian H, Aposhian M. Arsenic toxicology: five questions. Chem Res Toxicol. 2006 Jan;19(1):1–15.

[70] Percy AJ, Gailer J. Methylated trivalent arsenic-glutathione complexes are more stable than their arsenite analog. Bioinorg Chem Appl. 2008;2008:1–8.

[71] Hayakawa T, Kobayashi Y, Cui X, Hirano S. A new metabolic pathway of arsenite: arsenic–glutathione complexes are substrates for human arsenic methyltransferase Cyt19. Archives of toxicology. 2005;79(4):183–191.

[72] Thomas DJ, Li J, Waters SB, Xing W, Adair BM, Drobna Z, et al. Arsenic (+ 3 oxidation state) methyltransferase and the methylation of arsenicals. Experimental Biology and Medicine. 2007;232(1):3.

[73] Ord MG, Stocken LA. A contribution to chemical defence in World War II. Trends in biochemical sciences. 2000;25(5):253.

[74] Ahmed M, Ahuja S, Alauddin M, Hug S, Lloyd J, Pfaff A, et al. Ensuring safe drinking water in Bangladesh. Science. 2006;314(5806):1687.

[75] Rahman M, Chowdhury U, Mukherjee S. Chronic arsenic toxicity in Bangladesh and West Bengal, India—a review and commentary. Clinical Toxicology. 2001 Jan;39(7):683–700.

[76] Chaudhuri SD, Ghosh P, Sarma N, Majumdar P, Sau T, Basu S, et al. Genetic variants associated with arsenic susceptibility: study of purine nucleoside phosphorylase, arsenic (+ 3) methyltransferase, and glutathione s-transferase omega genes. Environmental Health Perspectives. 2008;116(4):501.

[77] Wang L, Zhou G, Liu P, Song J, Liang Y, Yan X, et al. Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia. Proceedings of the National Academy of Sciences. 2008;105(12):4826.

[78] Wang ZY, Chen Z. Acute promyelocytic leukemia: from highly fatal to highly curable. Blood. 2008;111(5):2505.

[79] Nasr R, Guillemin MC, Ferhi O, Soilihi H, Peres L, Berthier C, et al. Eradication of acute promyelocytic leukemia-initiating cells through PML-RARA degradation. Nature medicine. 2008;14(12):1333–1342.

[80] Kogan S. Poisonous Contacts. Science. 2010;328(5975):184.

[81] Pelicano H, Feng L, Zhou Y, Carew JS, Hileman EO, Plunkett W, et al. Inhibition of mitochondrial respiration. Journal of Biological Chemistry. 2003;278(39):37832.

[82] Wang J, Li L, Cang H, Shi G, Yi J. NADPH oxidase-derived reactive oxygen species are responsible for the high susceptibility to arsenic cytotoxicity in acute promyelocytic leukemia cells. Leuk Res. 2008 Mar;32(3):429–36.

[83] Lallemand-Breitenbach V, Jeanne M, Benhenda S, Nasr R, Lei M, Peres L, et al. Arsenic degrades PML or PML–RARα through a SUMO-triggered RNF4/ubiquitin-mediated pathway. Nature cell biology. 2008;10(5):547–555.

[84] Zhang XW, Yan XJ, Zhou ZR, Yang FF, Wu ZY, Sun HB, et al. Arsenic trioxide controls the fate of the PML-RARalpha oncoprotein by directly binding PML. Science. 2010 Apr;328(5975):240–3.

[85] Ito K, Bernardi R, Morotti A, Matsuoka S, Saglio G, Ikeda Y, et al. PML targeting eradicates quiescent leukaemia-initiating cells. Nature. 2008;453(7198):1072–1078.

[86] Haupt S, di Agostino S, Mizrahi I, Alsheich-Bartok O, Voorhoeve M, Damalas A, et al. Promyelocytic leukemia protein is required for gain of function by mutant p53. Cancer research. 2009;69(11):4818.

[87] Don AS, Kisker O, Dilda P, Donoghue N, Zhao X, Decollogne S, et al. A peptide trivalent arsenical inhibits tumor angiogenesis by perturbing mitochondrial function in angiogenic endothelial cells. Cancer Cell. 2003;3(5):497–509.

[88] Maignan S, Guilloteau J, Zhou-Liu Q, Clément-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases1. Journal of molecular biology. 1998;282(2):359–368.

[89] Li R, Haile JD, Kennelly PJ. An arsenate reductase from Synechocystis sp. strain PCC 6803 exhibits a novel combination of catalytic characteristics. Journal of bacteriology. 2003;185(23):6780.

[90] Mukhopadhyay R, Shi J, Rosen B. Purification and Characterization of Acr2p, the Saccharomyces cerevisiae Arsenate Reductase. Journal of Biological Chemistry. 2000;275(28):21149.

[91] Liu J, Rosen B. Ligand interactions of the ArsC arsenate reductase. Journal of Biological Chemistry. 1997;272(34):21084.

[92] Zegers I, Martins JC, Willem R, Wyns L, Messens J. Arsenate reductase from S. aureus plasmid pI258 is a phosphatase drafted for redox duty. Nat Struct Biol. 2001 Oct;8(10):843–7.

[93] Messens J, Martins JC, Belle KV, Brosens E, Desmyter A, Gieter MD, et al. All intermediates of the arsenate reductase mechanism, including an intramolecular dynamic disulfide cascade. Proc Natl Acad Sci USA. 2002 Jun;99(13):8506–11.

[94] Messens J, Hayburn G, Desmyter A, Laus G, Wyns L. The essential catalytic redox couple in arsenate reductase from Staphylococcus aureus. Biochemistry. 1999 Dec;38(51):16857–65.

[95] Martin P, DeMel S, Shi J, Gladysheva T, Gatti D, Rosen B, et al. Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. Structure. 2001;9(11):1071–1081.

[96] Messens J, Martins J, Brosens E, Belle KV, Jacobs D, Willem R, et al. Kinetics and active site dynamics of Staphylococcus aureus arsenate reductase. Journal of Biological Inorganic Chemistry. 2002;7(1):146–156.

[97] Denu JM, Dixon JE. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. Current opinion in chemical biology. 1998;2(5):633–641.

[98] Xia B, Vlamis-Gardikas A, Holmgren A, Wright PE, Dyson HJ. Solution structure of Escherichia coli glutaredoxin-2 shows similarity to mammalian glutathione-S-transferases1. Journal of molecular biology. 2001;310(4):907–918.

[99] Stevens SY, Hu W, Gladysheva T, Rosen BP, Zuiderweg ERP, Lee L. Secondary structure and fold homology of the ArsC protein from the Escherichia coli arsenic resistance plasmid R773. Biochemistry. 1999;38(31):10178–10186.

[100] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25(17):3389.

[101] DeMel S, Shi J, Martin P, Rosen BP, Edwards BFP. Arginine 60 in the ArsC arsenate reductase of E. coli plasmid R773 determines the chemical nature of the bound As (III) product. Protein Sci. 2004;13(9):2330–2340.

[102] Board P, Coggan M, Cappello J, Zhou H, Oakley A, Anders M. S-(4-Nitrophenacyl)glutathione is a specific substrate for glutathione transferase omega 1-1. Anal Biochem. 2008 Mar;374(1):25–30.

[103] Praznikar J, Afonine PV, Guncar G, Adams PD, Turk D. Averaged kick maps: less noise, more signal... and probably less bias. Acta Crystallographica Section D: Biological Crystallography. 2009;65(9):921–931.

[104] Axarli IA, Rigden DJ, Labrou NE. Characterization of the ligandin site of maize glutathione S-transferase I. Biochemical Journal. 2004;382(Pt 3):885.

[105] Reinemer P, Prade L, Hof P, Neuefeind T, Huber R, Zettl R, et al. Three-dimensional structure of glutathione S-transferase from Arabidopsis thaliana at 2.2 Å resolution: structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. Journal of molecular biology. 1996;255(2):289–309.

[106] Oakley AJ, Bello ML, Nuccetelli M, Mazzetti AP, Parker MW. The ligandin (non-substrate) binding site of human pi class glutathione transferase is located in the electrophile binding site (H-site) 1. Journal of molecular biology. 1999;291(4):913–926.

[107] McTigue MA, Williams DWR, Tainer JA. Crystal Structures of a Schistosomal Drug and Vaccine Target: Glutathione S-Transferase from Schistosoma japonica and its Complex with the Leading Antischistomal Drug Praziquantel. Journal of molecular biology. 1995;246(1):21–27.

[108] Ji X, Rosenvinge ECV, Johnson WW, Armstrong RN, Gilliland GL. Location of a potential transport binding site in a sigma class glutathione transferase by x-ray crystallography. Proceedings of the National Academy of Sciences. 1996;93(16):8208.

[109] Patskovska L, Fedorov A, Patskovsky Y, Almo S, Listowsky I. Expression, crystallization and preliminary X-ray analysis of ligand-free human glutathione S-transferase M2-2. Acta Crystallographica Section D: Biological Crystallography. 1998;54(3):458–460.

[110] Tsodikov OV, Jr MTR, Sergeev YV. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. Journal

of computational chemistry. 2002;23(6):600–609.

[111] Morel M, Ngadin AA, Droux M, Jacquot JP, Gelhaye E. The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete Phanerochaete chrysosporium. Cellular and molecular life sciences. 2009;66(23):3711–3725.

[112] Xun L, Belchik S, Xun R, Huang Y, Zhou H, Sanchez E, et al. S-glutathionyl-(chloro) hydroquinone reductases: a novel class of glutathione transferases. Biochem J. 2010;428:419–427.

[113] Meux E, Prosper P, Ngadin A, Didierjean C, Morel M, Dumarçay S, et al. Glutathione Transferases of Phanerochaete chrysosporium. Journal of Biological Chemistry. 2011;286(11):9162.

[114] Garcerá A, Barreto L, Piedrafita L, Tamarit J, Herrero E. Saccharomyces cerevisiae cells have three Omega class glutathione S-transferases acting as 1-Cys thiol transferases. Biochemical Journal. 2006;398(Pt 2):187.

[115] Ji X, Johnson WW, Sesay MA, Dickert L, Prasad SM, Ammon HL, et al. Structure and function of the xenobiotic substrate binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene. Biochemistry. 1994 Feb;33(5):1043–52.

[116] Ji X, Johnson WW, Sesay MA, Dickert L, Prasad SM, Ammon HL, et al. Structure and function of the xenobiotic substrate binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9, 10-dihydrophenanthrene. Biochemistry. 1994;33(5):1043–1052.

[117] Prade L, Huber R, Manoharan TH, Fahl WE, Reuter W. Structures of class pi glutathione S-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. Structure. 1997;5(10):1287–1295.

[118] Griffin WST, Mrak RE. Interleukin-1 in the genesis and progression of and risk for development of neuronal degeneration in Alzheimer's disease. Journal of leukocyte biology. 2002;72(2):233.

[119] Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, et al. Age at onset in two common neurodegenerative diseases is genetically controlled. The American Journal of Human Genetics. 2002;70(4):985–993.

[120] Ledford H. Fever pitch. Nature. 2007;450:600–601.

[121] Braddock M, Quinn A. Targeting IL-1 in inflammatory disease: new opportunities for therapeutic intervention. Nat Rev Drug Discov. 2004 Apr;3(4):330–9.

[122] Eder C. Mechanisms of interleukin-1 [beta] release. Immunobiology. 2009;214(7):543–553.

[123] Perregaux DG, Bhavsar K, Contillo L, Shi J, Gabel CA. Antimicrobial peptides initiate IL-1β posttranslational processing: a novel role beyond innate immunity. The Journal of Immunology. 2002;168(6):3024.

[124] Ozturk A, Desai P, Minster R, Dekosky S, Kamboh M. Three SNPs in the GSTO1, GSTO2 and PRSS11 genes on chromosome 10 are not associated with age-at-onset of Alzheimer's disease. Neurobiol Aging. 2005 Aug;26(8):1161–5.

[125] Whitbread A, Mellick G, Silburn P, Couteur DL, Board P. Glutathione transferase Omega class polymorphisms in Parkinson disease. Neurology. 2004;62(10):1910.

[126] Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, Deng SJJ, et al. Crystallization of protein-ligand complexes. Acta Crystallographica Section D: Biological Crystallography. 2006;63(1):72–79.

[127] Perregaux D, McNiff P, Laliberte R, Hawryluk N, Peurano H, Stam E, et al. Identification and characterization of a novel class of interleukin-1 post-translational processing inhibitors. Journal of Pharmacology and Experimental Therapeutics. 2001;299(1):187.

[128] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of Applied Crystallography. 2007;40(4):658–674.

[129] Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallographica Section D: Biological Crystallography. 2004;60(12):2184–2195.

[130] Guan X, Hoffman B, McFarland D, Gilkerson K, Dwivedi C, Erickson A, et al. Glutathione and mercapturic acid conjugates of sulofenur and their activity against a human colon cancer cell line. Drug Metabolism and Disposition. 2002;30(3):331.

[131] Cameron AD, Sinning I, L'Hermite G, Olin B, Board PG, Mannervik B, et al. Structural analysis of human alpha-class glutathione transferase A1-1 in the apo-form and in complexes with ethacrynic acid and its glutathione conjugate. Structure. 1995;3(7):717–727.

[132] Otting G. Prospects for lanthanides in structural biology by NMR. Journal of biomolecular NMR. 2008;42(1):1–9.

[133] Jia X, Yagi H, Su X, Stanton-Cook M, Huber T. Engineering [Ln (DPA) 3] 3- binding sites in proteins: a widely applicable method for tagging proteins with lanthanide ions. Journal of biomolecular NMR. 2011;50:411–420.

[134] Zhou H, Brock J, Casarotto MG, Oakley AJ, Board PG. Novel folding and stability defects cause a deficiency of human glutathione transferase omega 1. Journal of Biological Chemistry. 2011;286(6):4271.

Novel Folding and Stability Defects Cause a Deficiency of Human Glutathione Transferase Omega 1

Myself and colleagues have recently published the crystal structure of the E155/E208 variant of this polymorphic enzyme. The coordinates have been submitted to the Protein Data Bank (PDB), deposition code 3LFL. Published in conjunction with the results of stability and unfolding experiments, our findings suggest that the enzyme possesses an increased propensity to unfold and once partially denatured, has an impaired ability to refold again. This agreed with the crystallographic data showing that the deletion of glutamate 155 results in the premature termination of helix α5, concurrently increasing flexibility of the N-terminal domain and the amount of hydrophobic residues at the solvent interface [134].

Although the preliminary coordinates of this structure were submitted as part of my honours thesis, subsequent refinement and interpretation that made this paper possible was performed during my PhD candidature.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 6, pp. 4271–4279, February 11, 2011

Novel Folding and Stability Defects Cause a Deficiency of Human Glutathione Transferase Omega 1*

Received for publication, October 25, 2010, and in revised form, November 17, 2010 Published, JBC Papers in Press, November 24, 2010, DOI 10.1074/jbc.M110.197822

Huina Zhou[‡], **Joseph Brock**[§], **Marco G. Casarotto**[‡], **Aaron J. Oakley**[§], **and Philip G. Board**[‡1]

From the [‡]*John Curtin School of Medical Research and the* [§]*Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 2601, Australia*

The polymorphic deletion of Glu-155 from human glutathione transferase omega1 (GSTO1-1) occurs in most populations. Although the recombinant ΔGlu-155 enzyme expressed in *Escherichia coli* is active, the deletion causes a deficiency of the active enzyme *in vivo*. The crystal structure and the folding/unfolding kinetics of the ΔGlu-155 variant were determined in order to investigate the cause of the rapid loss of the enzyme in human cells. The crystal structure revealed altered packing around the Glu-155 deletion, an increase in the predicted solvent-accessible area and a corresponding reduction in the buried surface area. This increase in solvent accessibility was consistent with an elevated Stern-Volmer constant. The unfolding of both the wild type and ΔGlu-155 enzyme in urea is best described by a three-state model, and there is evidence for the more pronounced population of an intermediate state by the ΔGlu-155 enzymes. Studies using intrinsic fluorescence revealed a free energy change around 14.4 kcal/mol for the wild type compared with around 8.6 kcal/mol for the ΔGlu-155 variant, which indicates a decrease in stability associated with the Glu-155 deletion. Urea induced unfolding of the wild type GSTO1-1 was reversible through an initial fast phase followed by a second slow phase. In contrast, the ΔGlu-155 variant lacks the slow phase, indicating a refolding defect. It is possible that in some conditions *in vivo*, the increased solvent-accessible area and the low stability of the ΔGlu-155 variant may promote its unfolding, whereas the refolding defect limits its refolding, resulting in GSTO1-1 deficiency.

The omega class glutathione transferases (GSTO)[2] are among the more recently identified members of the cytosolic glutathione transferase super family (1, 2). Members of the omega class have a number of novel features and have been implicated in important metabolic pathways and in the etiology of some common neurological diseases (3–7). In contrast to the other human GSTs, which have active site tyrosine or serine residues, the omega class GSTs have an active site cysteine residue that is essential for its primary catalytic activities (1). The omega class GSTs were originally shown to catalyze thioltransferase and ascorbate reductase reactions, and other studies have shown that they also catalyze the reduction of monomethyl arsonate in the methylation pathway for the disposition of arsenic (3, 4). In addition, recent studies have shown that GSTO1-1 has *S*-phenacyl glutathione reductase activity and plays a significant role in the biotransformation of α-haloketones (8, 9). As well as these catalytic activities, GSTO1-1 has been implicated in the post-translational processing and activation of the proinflammatory mediator interleukin-1β (10) and has also been found to act as an inhibitor of cardiac muscle ryanodine receptor Ca^{2+} channels (11). Interestingly, some genetic mapping studies have strongly implicated an omega class GST as a factor influencing the age at onset of Alzheimer and Parkinson diseases (5, 6, 12). Other studies have implicated genetic variation in *GSTO1* in vascular dementia, stroke, and cerebrovascular atherosclerosis (7, 13). The mechanism by which genetic variation mediates these effects has not been determined, but it could be via the effects on the activity or expression of the omega class GSTs.

Previous studies have identified a range of polymorphisms in the GSTO genes (2, 3, 14). Although a number of single nucleotide polymorphisms have been identified in non-coding regions, their effects on transcription, splicing, and mRNA stability have yet to be evaluated (15). In contrast, several polymorphisms in coding regions have been studied in more detail (2, 3, 14, 16). A polymorphic deletion of three base pairs at the junction of exon 4 and intron 4 is of particular interest. This deletion could result in the deletion of Glu-155 or a splicing abnormality. This polymorphism has been identified in a range of different ethnic groups reaching a frequency of >10% in one Chinese sample (2, 14, 17–19). The deletion occurs in two haplotypes characterized by a Glu or Lys at position 208 (19). Recombinant expression of the Glu-155 deletion variant in *Escherichia coli* affords an active enzyme that has decreased heat stability (2, 19). We recently reported that the T-47D breast cancer cell line is hemizygous for the Glu-155 deletion allele and is deficient in GSTO1-1 enzymatic activity (20). In addition, lymphoblastoid cell lines from individuals who were heterozygous for the Glu-155 deletion were reported to have only 50% of normal catalytic activity with 4-nitrophenacylglutathione, a specific GSTO1-1 substrate (20). These studies suggest that the Glu-155 deletion variant is very compromised *in vivo*.

Given the role played by GSTO1-1 in the disposition of α-haloketones (8) and the targeting of GSTO1-1 by anti-in-

* This work was supported by National Health and Medical Research Council Project Grant 366731.

The atomic coordinates and structure factors (code 3LFL) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] To whom correspondence should be addressed: Curtin School of Medical Research, Australian National University, P. O. Box 334, Canberra, ACT 2601, Australia. Tel.: 61-2-61254714; Fax: 61-2-61254712; E-mail: Philip.Board@anu.edu.au.

[2] The abbreviations used are: GSTO, omega class glutathione transferase; ANS, 1-anilino-8-naphthalenesulfonate.

flammatory drugs (10) as well as the proposed role of genetic variation in *GSTO1* in the progress of several neurological disorders (6, 7), it is apparent that natural polymorphisms in *GSTO1* that affect function are of significant pharmacogenetic and clinical interest. Consequently we have investigated the effect of the polymorphic Glu-155 deletion on the enzyme structure and stability by solving its crystal structure and studying its unfolding equilibrium.

EXPERIMENTAL PROCEDURES

Expression of Recombinant GSTO1-1 Variants—In this study several recombinant GSTO1-1 isoenzymes have been studied. There is a common polymorphism where Asp-140 substitutes Ala-140. This substitution does not appear to influence function, but because Ala-140 is the most common allele, we refer to it as the wild type and the substituted variant as A140D. In each case the Glu-155 deletion variants have alanine at residue 140. cDNAs encoding GSTO1 wild type, GSTO1-A140D, GSTO1 ΔGlu-155/Glu-208, and GSTO1 ΔGlu-155/Lys-208 were cloned into the pHUE expression vector and transfected into *E. coli* BL21(DE3). The expression and purification of recombinant proteins using the pHUE vector and ubiquitin cleavage techniques have been previously described in detail (21). Briefly, the recombinant His_6-ubiquitin-GSTO fusion protein was purified by chromatography on nickel-agarose. The His-ubiquitin fusion moiety was removed by digestion with the catalytic core of Usp2 and rechromatography on nickel-agarose. The purified proteins were dialyzed against 20 mM Tris, pH 8, 60 mM NaCl, 5 mM dithiothreitol for storage and crystallization experiments.

Crystallization and Structural Determination of GSTO1-1 ΔGlu-155/Glu-208—Unsuccessful attempts were made to grow crystals of GSTO1 ΔGlu-155 under identical conditions to those reported for the native enzyme (1). The protein was, therefore, subjected to the index screen (Hampton Research) to find new conditions. The vapor diffusion technique was employed using a 96-well sitting drop crystallization tray (Nunc, Roskilde, Denmark). For screening, 1 μl of protein (12 mg/ml) was mixed with 1 μl of 10 mM glutathione and 1 μl of reservoir solution. Reservoirs contained 100 μl of each screen condition. The tray was then sealed with Crystal Clear tape (Hampton Research) and allowed to equilibrate at 18 °C. After 6 days, crystals were observed in condition 62 (0.1 M Tris, pH 8.5, 0.2 M trimethyl amine *N*-oxide dihydrate, 20% (w/v) polyethylene glycol monomethyl ether 2000). Scaling up was achieved via a doubling of the overall crystallization volume using the same method. Crystals were flash-frozen at 100 K using an Oxford cryostream. Cryoprotection was by sequential transfer to artificial mother liquor containing 20% glycerol by 5% increments. The crystals were then transferred to liquid nitrogen and stored until subsequent data collection using a Rigaku RU-200 rotating anode x-ray generator with mirror optics and Mar345 desktop beamline. Processing of data collected from a single crystal was performed using the programs DENZO and SCALEPACK (22). Phase information was obtained by molecular replacement using the structure of the native enzyme (1) as a search model with the program MOLREP (23). Model rebuilding to match experimental density, including the deletion of Glu-155, was achieved with the programs O (24) and Coot (25) and alternated with restrained refinement by the program REFMAC (26). The program PHENIX (27) was then employed for the final stages of refinement using the simulated annealing option with default settings. Finally, the program MolProbity (28) was used to add hydrogens to the structure and diagnose Asn/Gln/His flips before validation deposition (PDB code 3LFL).

Spectroscopy and Activity Measurements—All fluorescence data were measured on PerkinElmer Life Sciences LS 50B luminescence spectrophotometer. Tryptophan fluorescence was used to monitor structural change. Excitation was at 295 nm, and fluorescence was recorded in the range of 305–430 nm. 1-Anilino-8-naphthalenesulfonate (ANS) was added to protein samples in a molar ratio of 50:1 (29) and equilibrated for 50 min for complete interaction. ANS binding spectra were recorded in the range of 405–540 nm with excitation at 390 nm. Enzymatic activity was determined with the GSTO1-1 specific substrate *S*-(4-nitrophenacyl)-glutathione as described previously (9).

Tryptophan Fluorescence Quenching by Acrylamide—Acrylamide (8 M) was added to 1 μM GSTO1 proteins (calculated by a molar extinction coefficient of 82,800 $M^{-1}cm^{-1}$ at 280 nm) to achieve different concentrations and mixed thoroughly. Fluorescence intensities were monitored at 336 nm. Each protein sample was repeated at least three times and then averaged. Quenching data were analyzed by the Stern-Volmer equation (30),

$$\frac{F_o}{F} - 1 = K_{sv}[Q] \qquad \text{(Eq. 1)}$$

where F_o and F are the fluorescence intensities before and after the addition of acrylamide, $[Q]$ is the acrylamide concentration, and K_{SV} is the Stern-Volmer constant, which can reflect the burial of tryptophan residues and, thus, indicate the protein accessibility.

Circular Dichroism Measurement—Circular dichroism (CD) was monitored on Chirascan™ Circular Dichroism Spectrometer (Applied Photophysics) using a cuvette with a 0.1-cm light path. Ellipticity at 222 nm was collected and plotted against variation of urea concentration.

Equilibrium Unfolding Experiments—Urea-induced unfolding experiments were performed in 20 mM sodium dihydrogen phosphate with 40 mM ammonium phosphate, pH 7.0, at 1 μM protein concentration. To reach equilibrium, protein samples with different urea concentrations (0–9 M) were incubated at 30 °C for 5 h and then cooled down to room temperature for intrinsic fluorescence, ANS binding, and circular dichroism measurements. For the analysis of the intrinsic fluorescence the ratio of emission intensity at 333 nm (folded protein) and 355 nm (unfolded protein) was determined, and the relative intensity at 475 nm was determined for ANS binding. Data from three-four separate replicates from two purification batches of each protein were averaged.

Urea unfolding curves were fitted to a two-state ($N_2 \leftrightarrow 2U$) unfolding model or a three-state model ($N_2 \leftrightarrow I_2 \leftrightarrow 2U$) (31, 32) using SigmaPlot Version 10.0. The equilibrium fractions

TABLE 1

X-ray data collection and refinement statistics

r.m.s.d., root mean square deviation.

Space group	$P2_12_12$
Unit cell parameters (Å)	$a = 73.0, b = 201.9, c = 53.4$
Reflections measured/unique	389,373/44,498
Resolution range (Å)	36.51-2.10 (2.18-2.10)[a]
R-merge (%)	12.8 (37.0)[a]
Completeness	94.4 (80.7)[a]
Mean $I/\sigma I$	21.4 (3.5)[a]
R/R_{free} (%)	19.91/26.05
r.m.s.d. bonds (Å)	0.008
r.m.s.d. angles	1.120°
PDB accession code	3LFL

[a] The highest resolution bin.

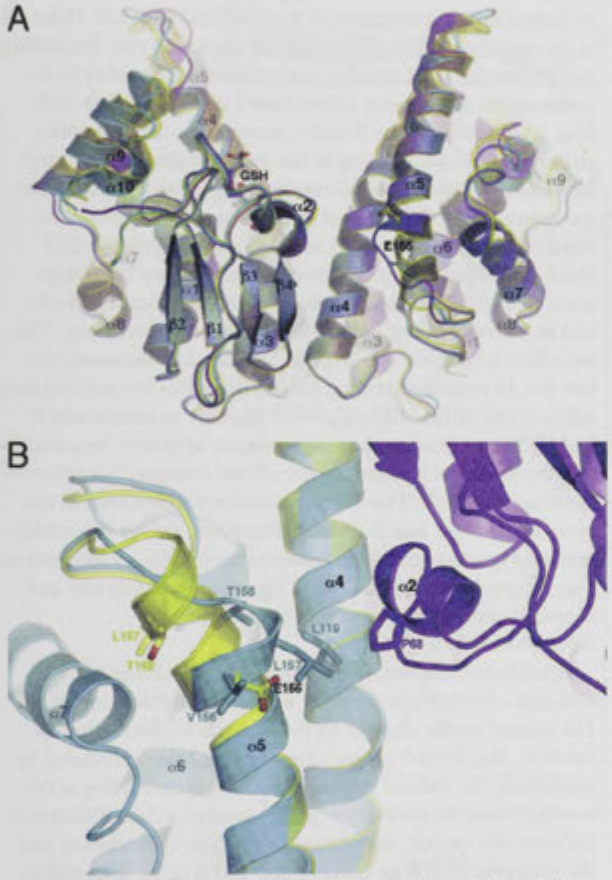


FIGURE 1. **Structural changes induced by the deletion of Glu-155.** *A*, a view perpendicular to the crystallographic 2-fold axis of the native structure (translucent *yellow*), the ΔGlu-155 A monomer (*cyan*), and the non-crystallographic 2-fold axis of the ΔGlu-155 B and C monomers (*magenta*), which produce the GSTO1-1 physiological dimer. The position of Glu-155 and GSH is shown in stick representation. *B*, a closer inspection of the deletion site within the ΔGlu-155 A-monomer (*cyan*) compared with the wild type enzyme (*yellow*) is shown. Changes to the associated residue side chain conformations are in stick representation. The new interaction formed at the dimer interface is also visible. The adjacent monomer is shown in *magenta*. The figures were generated using PyMol.

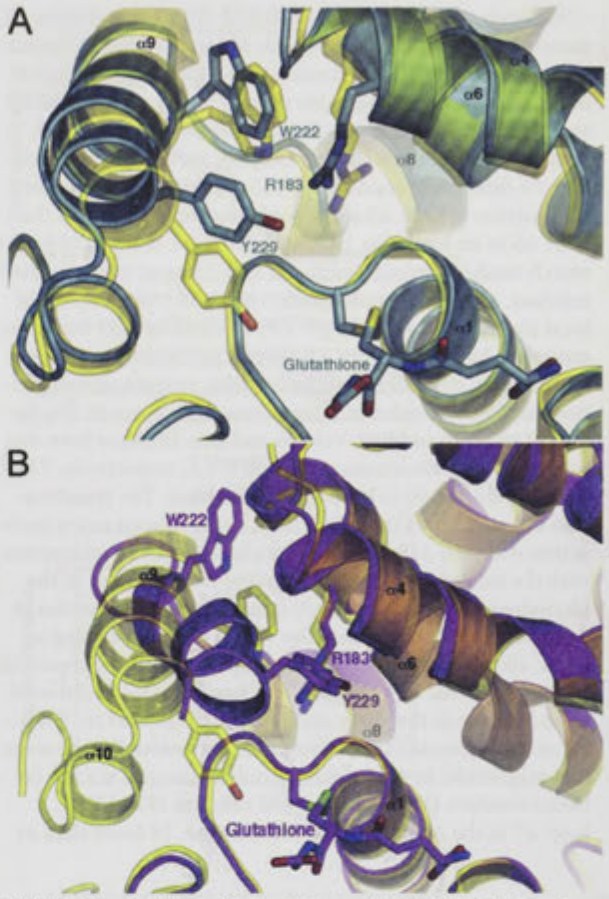


FIGURE 2. **The flexibility of helix α9 observed in the ΔGlu-155 structure.** *A*, a closer view of the A monomer active site reveals several key differences compared with the native enzyme. Of most interest is the different rotamer adopted by Trp-222, dramatically increasing the hydrophobicity within the conserved H-site. The ΔGlu-155 A-monomer is shown in *cyan* compared with the wild type enzyme in *yellow*. *B*, comparison of the native structure to that of the B monomer reveals much more pronounced structural change. Helix α9 has shifted toward the site of glutathione binding by several angstroms, dramatically reducing the size of the H-site. The ΔGlu-155 B-monomer is shown in *magenta* with the overlaid native structure in *yellow*. The figures were generated using PyMol.

of native, intermediate, and denatured state were calculated by a previously described method (32).

Protein Refolding—Refolding of unfolded proteins was detected using *S*-(4-nitrophenacyl)-glutathione reduction assay and tryptophan fluorescence recovery. For activity reversibility, proteins were brought to a concentration of 1 μM in 6 M urea and allowed to unfold for 5 h at 30 °C. The proteins were then diluted 6 times and allowed to refold at room temperature. Enzyme activity was determined after 1 h. When refolding was monitored by intrinsic fluorescence, 30 μM proteins were fully unfolded in 7.4 M urea. The enzymes were then diluted 10 times, and the emission intensities were recorded every minute. Curve fitting was undertaken in SigmaPlot Version 10.0.

RESULTS

Crystal Structure of GSTO1 ΔGlu-155/Lys-208—To determine whether the deletion of Glu-155 caused any significant structural alterations, we crystallized and solved the structure of the GSTO1 ΔGlu-155/Glu-208 variant. The crystals were found to be of a different space group ($P2_12_12$) compared with the native enzyme ($P3_121$) and diffracted to 2.1 Å. Table 1 summarizes the crystallographic data and model statistics.

Molecular replacement in MOLREP (23) revealed three monomers in the asymmetric unit. Monomer A sits adjacent to a crystallographic 2-fold axis so as to form a physiological dimer with its crystallographic partner (A-A). Monomers B and C form a second dimer (B-C). The overall topology is similar to that of the wild type enzyme, except for the following; the deletion of residue Glu-155 results in the premature termination of helix α5 and a conformational change in the helix α5 to α6 loop (Fig. 1*A*). Its subsequent removal from the search model required significant rebuilding of the associated residues, a process coupled with restrained refinement until local aberrations in the $mF_o - DF_c$ electron density map were removed. As shown in Fig. 1*B*, several perturbations in the surrounding structure contribute to this enzyme altered phenotype. The hydrophobic interactions (not shown in this figure) of adjacent residues Val-156 and Leu-157 have been disrupted by Cα translations of 5.3 and 8.0 Å, respectively. This has caused Val-156 to be exposed to solvent. The translocation of Leu-157, on the other hand, has produced a new interaction with Leu-119 and, of particular interest, an interaction with the side chain of Pro-68 from the other subunit in the physiological dimer, effectively creating a new interaction at dimer interface. In addition, the Thr-158 Cα has shifted by 6.4 Å, allowing its side chain to occupy the position previously inhabited by that of Val-156. Such changes have contributed to an increase in the water accessibility of Leu-157 to >100 Å^2 and an increase in the buried area at the interface of twice this magnitude. In total there are only 7 contacts ≤3.5 Å between residues 150–164 of α5 and residues 192–202 of loop-α7 in the ΔGlu-155/Glu-208 enzyme, 18 fewer than in the wild type enzyme. The altered packing has likely played a role in the observed refolding anomaly of Glu-155 deletion variants discussed below. The alterations in the dimer interface do not appear to preclude the formation of heterodimers between wild type and deletion-containing monomers.

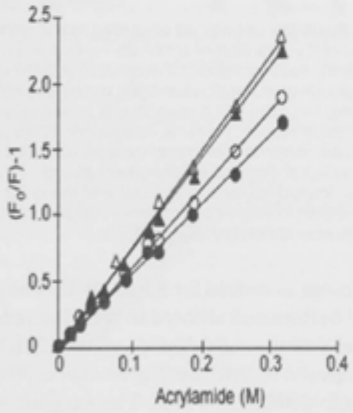


FIGURE 3. **Tryptophan fluorescence quenching by acrylamide.** The experimental values and trend lines are shown as wild type (●), A140D (○), ΔGlu-155/Glu-208 (▼), and ΔGlu-155/Lys-208 (△).

The potential flexibility of the C-terminal region can be observed in the ΔGlu-155/Glu-208 structure. Comparison of the active site of monomer A with the wild type enzyme reveals a rotation of the putative "H-site" residue, Trp-222, resulting in the removal of its indole nitrogen from the binding pocket, increasing its hydrophobic character. As shown in Fig. 2*A*, the entrance to the binding pocket has also been partially occluded by the movement of Arg-183 and Tyr-229. Helix α9, of monomer A, which forms part of the active site, has shifted away from the GSH binding pocket (G-site), possibly to accommodate movements in residues Tyr-229 and Trp-222 (Fig. 2*A*). Although the B and C monomers display similar structural changes relating to the deletion site, it is the conformational shift of their C termini that is most remarkable. As shown in Fig. 2*B* helix α9 of monomer B has moved toward the G-site and rotated so that Leu-228 and Tyr-229 block the H-site entrance. There is a significant rearrangement of Trp-222, which has been excluded from the H-site and is stacked against helix α6, exposed to bulk solvent. The net effect is to make the "H-site" smaller. Furthermore, the last 10–14 residues forming helix α10 and its connecting loop fail to show substantial $mF_o - DF_c$ density in monomers B and C. These significant changes do not appear to be a consequence of crystal packing, as the crystal contacts in each monomer are different. The different conformations seen in the two dimers (A-A and B-C) may, therefore, reflect the conformational space sampled in solution. Such helix α9 movements may be crucial to the binding of ligands of variable size and shape, for example monomethyl arsonate and dehydroascorbate.

Equilibrium Unfolding/Folding of GSTO1 ΔGlu-155/Glu-208—Initially, to determine whether the deletion of Glu-155 caused subtle changes to the enzyme conformation in solution, the Stern-Volmer constant (K_{SV}) was calculated by quenching the natural tryptophan fluorescence using acrylamide. Linear Stern-Volmer plots revealed a clear difference between the variant enzymes with the Glu-155 deletion and the enzymes with A or D at residue 140 (Fig. 3). This difference is reflected numerically in K_{SV} for each enzyme (5.258 (R^2 = 0.9989) for wild type (Ala-140), 5.840 (R^2 = 0.9982) for A140D, 6.975 (R^2 = 0.999) for ΔGlu-155/Glu-208, and 7.299 (R^2 = 0.9979) for ΔGlu-155/Lys-208). The data suggest that the deletion of Glu-155 increases the acces-

TABLE 2

Interface and surface area (SA) of ΔGlu-155/Glu-208 and wild type crystal structure

All calculations come from online PISA (35) of EBI services.

	Solved residues	SA of dimer	Total buried area of dimer	Total SA of Trp residues in dimer	Buried area of dimer interface
		Å^2	Å^2	Å^2	Å^2
GSTO1 wild type (PDB code 1EEM)	Ser-5–Leu-241	20,961.5	3,859.9	32.48	1,885.6
ΔGlu-155 monomer A	Ser-5–Leu-240	21,639.1	3,317.3	33.74	2,116.6

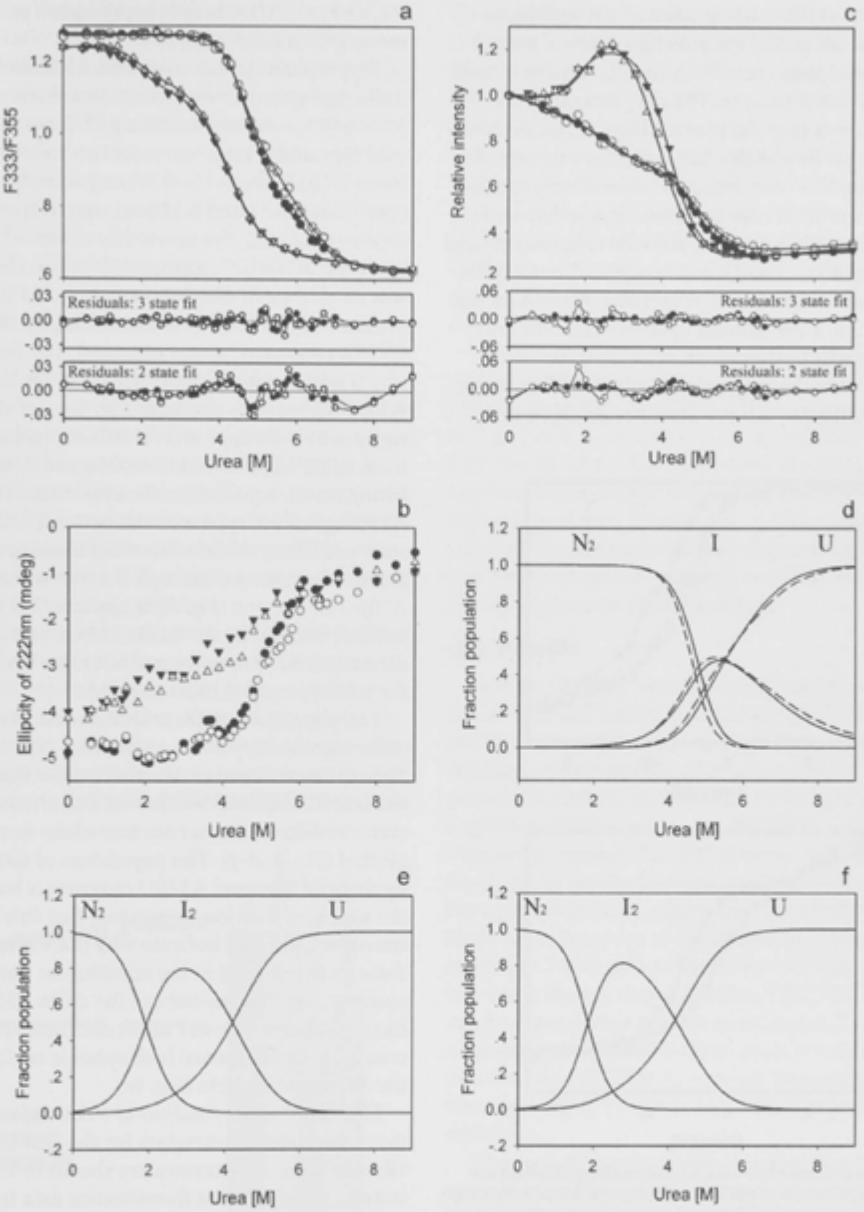


FIGURE 4. **The equilibrium unfolding of GSTO1-1 in urea.** Unfolding was monitored by tryptophan fluorescence expressed as the ratio of fluorescence at 333 nm to the fluorescence at 355 nm (*a*), ellipticity at 222 nm of circular dichroism (*b*), and relative intensity of ANS binding (*c*). The three state curve fits for $N_2 \leftrightarrow I_2 \leftrightarrow 2U$ are shown in *a* and *c*, and the *panels immediately below a* and *c* show the residuals obtained after fitting the data to either two or three state models. The calculated population of the native, intermediate, and unfolded states of GSTO1-1 as a function of urea concentration are plotted for wild type and A140D in *panel d* (*solid line* for wild type and *dashed line* for A140D), ΔGlu-155/Glu-208 in *panel e*, and ΔGlu-155/Lys-208 in *panel f*. The experimental values for each enzyme are shown as wild type (●), A140D (○), ΔGlu-155/Glu-208 (▼), and ΔGlu-155/Lys-208 (△).

sibility of tryptophan residues. This is supported by calculations from the crystal structure of increased accessible surface area of total protein and of tryptophan residues in the ΔGlu-155 (Table 2).

Characterization of Urea-induced Unfolding—The unfolding of the GSTO1-1 enzymes in the presence of urea was examined using tryptophan fluorescence, CD, and ANS binding (Fig. 4). The fluorescence data (Fig. 4*a*) for the wild type and A140D enzymes indicate that they do not begin unfolding before the urea concentration exceeds 3.7 M and by visual inspection appear to exhibit a single transition phase. In contrast, the equilibrium denaturation of the ΔGlu-155 enzymes start to show an unfolding signal below 1 M urea and clearly have a biphasic transition profile. The difference in the unfolding profiles between the Glu-155-containing and ΔGlu-155 enzymes is also evident in the CD data (Fig. 4*b*) where the ΔGlu-155 enzymes start losing their helical secondary structure content at the lowest concentrations of urea. It is also of

interest to note that at the starting point of the unfolding curve in Fig. 4*b* the ΔGlu-155 variants have a lower helical content than the wild type enzymes, which is consistent with the ΔGlu-155/Glu-208 structure. The data from the ANS binding measurements (Fig. 4*c*) provide the clearest evidence that the unfolding profiles of the ΔGlu-155 enzymes are different from the Glu-155-containing enzymes. Using tryptophan fluorescence, protein concentration dependence was observed for the transition phase of the wild type enzyme and notably only during the second transition phase of the ΔGlu-155/Glu-208 enzyme (Fig. 5). Taken together, these data suggested that the equilibrium unfolding of the wild type and A140D enzymes might be described by a two-state model ($N_2 \leftrightarrow 2U$), whereas the unfolding of the ΔGlu-155 enzymes might best be described by a three-state unfolding model ($N_2 \leftrightarrow I_2 \leftrightarrow 2U$), where I_2 represents a partially unfolded dimeric intermediate (33, 34).

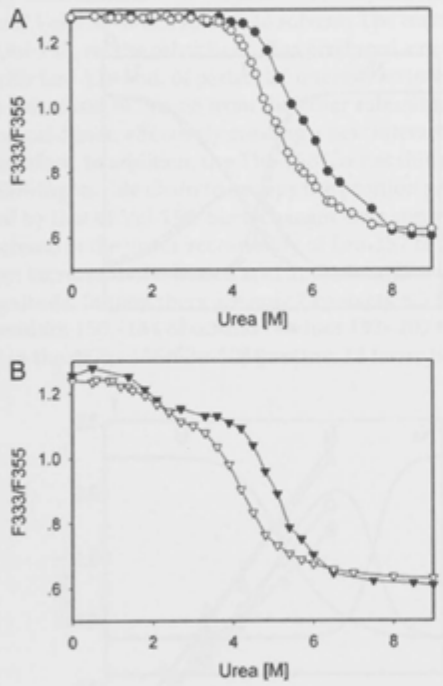


FIGURE 5. **The protein concentration dependence of equilibrium unfolding of GSTO1-1 variants in urea.** The *unfilled symbols* represent determinations at 1 μM protein, and the filled symbols represent determinations at 8 μM protein. The *circles* in *panel A* represent the wild type enzyme, and the *triangles* in *panel B* represent the ΔGlu-155/Glu-208 variant.

Tryptophan fluorescence and ANS binding data from the ΔGlu-155 enzymes were fitted to a three-state model with an $R^2 > 0.999$, whereas the fitting of fluorescence data from the wild type and A140D enzymes to a two-state model gave a lower R^2 (0.9986 and 0.9974, respectively) and lower free energy values (6.41 and 5.15 kcal/mol, respectively) that were inconsistent with the previously observed lower thermal stability for ΔGlu-155 enzymes (2). When the three-state model was used to re-fit the data from the wild type and A140D enzymes, a better fit (Fig. 4*a*) with a higher R^2 (0.9996 and 0.9993, respectively) was obtained. The distribution of residuals for a two-state and three-state fit for the wild type and A140D enzymes is shown in Fig. 4*a*, and the residual sums of squares for wild type and A140D enzymes were also improved from 0.0037 and 0.0062 to 0.0010 and 0.0017. Based on the fitting result, equilibrium denaturation of the wild type and A140D proteins appears to be better described by a three-state unfolding model. Given the absence of protein concentration dependence through the first transition phase of the ΔGlu-155 enzyme (Fig. 5), it appears that the transition intermediate formed by the ΔGlu-155 enzymes is likely to be a dimer (33, 34). However it is not clear whether this is the case for wild type and A140D enzymes.

The population of the transition intermediate appears to differ significantly between the Glu-155 containing and ΔGlu-155 enzymes. Based on the three-state model, the fractions of the transitional intermediate and the native and unfolded states at different urea concentrations were calculated and plotted (Fig. 4, *d–f*). The population of intermediate formed by the wild type and A140D enzymes is largely overlapped by the unfolded fraction, suggesting that this intermediate is very transitory and may indicate why there is no obvious intermediate state reflected in the equilibrium unfolding curves. In contrast, the intermediate of the ΔGlu-155 enzymes is clearly distinguishable at lower urea concentrations and has an increased area of exposed hydrophobic surface as illustrated by the ANS binding data (Fig. 4*c*).

Thermodynamic Analysis of Urea-induced Unfolding—The thermodynamic parameters for the Glu-155 containing and Glu-155 deletion enzymes are shown in Table 3. The three-state fit of tryptophan fluorescence data from the wild type and A140D enzymes provided a similar total free energy change of 14.4 and 14.3 kcal/mol, respectively, which is larger

TABLE 3

Thermodynamic parameters characterizing the urea-induced unfolding transition monitored by tryptophan fluorescence and ANS binding

The data were fitted to a three-state transition model via a dimeric intermediate.

	$\Delta G^{o}_{N\rightarrow I}$	$m_{N\rightarrow I}$	$[Urea]_{1/2,I}$	$\Delta G^{o}_{I\rightarrow U}$	$m_{I\rightarrow U}$	$[Urea]_{1/2,2}$	$\Delta G^{o}_{N\rightarrow U}$
	kcal/mol	*kcal/mol/M*	*M*	*kcal/mol*	*kcal/mol/M*	*M*	*kcal/mol*
Tryptophan fluorescence							
WT	8.51 ± 0.46	1.81 ± 0.11	4.69 ± 0.06	5.91 ± 2.42	1.00 ± 0.36	5.91 ± 0.33	14.42
A140D	9.16 ± 0.82	2.02 ± 0.20	4.55 ± 0.09	5.14 ± 1.67	0.87 ± 0.26	5.94 ± 0.22	14.3
ΔGlu-155/Glu-208	2.84 ± 0.79	1.48 ± 0.37	1.92 ± 0.12	5.78 ± 0.35	1.37 ± 0.08	4.22 ± 0.03	8.62
ΔGlu-155/Lys-208	3.34 ± 0.97	1.80 ± 0.49	1.86 ± 0.11	5.39 ± 0.31	1.29 ± 0.07	4.19 ± 0.03	8.73
ANS binding							
WT	7.65 ± 2.43	2.45 ± 0.80	3.11 ± 0.11	7.07 ± 0.42	1.44 ± 0.08	4.91 ± 0.03	14.72
A140D	5.43 ± 2.41	2.10 ± 0.91	2.58 ± 0.14	5.58 ± 0.57	1.13 ± 0.12	4.93 ± 0.06	11.01
ΔGlu-155/Glu-208	3.51 ± 1.55	1.95 ± 0.77	1.77 ± 0.17	6.54 ± 0.58	1.55 ± 0.13	4.20 ± 0.04	10.05
ΔGlu-155/Lys-208	3.89 ± 1.64	1.88 ± 0.79	2.06 ± 0.15	4.98 ± 0.62	1.27 ± 0.14	3.92 ± 0.07	8.87

than that obtained for the ΔGlu-155 enzymes (8.6 and 8.7 kcal/mol for ΔGlu-155/Glu-208 and ΔGlu-155/Lys-208, respectively). Similar results were obtained from the ANS binding data. The main difference between the wild type and the ΔGlu-155 variant occurs in the initial unfolding to the transition intermediate. Studies using intrinsic fluorescence revealed a free energy change for this first transition of around 8.5 kcal/mol for the wild type compared with around 2.8 kcal/mol for the ΔGlu-155 variant. These data suggest that the enzymes with the Glu-155 deletion are less stable than Glu-155-containing enzymes, and the lower stability of ΔGlu-155 enzymes comes primarily from a propensity to enter the first transition phase.

Reversibility of Unfolding—The ability of the enzymes to re-fold after urea-induced unfolding was evaluated by fluorescence recovery and the restoration of catalytic activity (Fig. 6). The two methods used here to monitor the refolding gave similar results. Refolding of the wild type Glu-155-containing enzymes was essentially complete after 1 h with the recovery of around 95%. In contrast, the enzymes with the Glu-155 deletion regained only around 70% of their original activity or fluorescence intensity over this period.

To identify whether the lower recovery of Glu-155 deletion enzyme came from a lower refolding rate, an exponential regression analysis was applied to the fluorescence recovery data. For the wild type and A140D enzymes, a two-phase exponential regression was better than a single-phase exponential regression as improvements in the fit were observed in R^2 (0.9984 and 0.9992 compared with 0.9924 and 0.9900, respectively) in the S.E. of estimates (0.0050 and 0.0069 compared with 0.0135 and 0.0158, respectively) and in the residual sums of squares (0.0005 and 0.0002 compared with 0.0022 and 0.0030, respectively). Interestingly the refolding data for the ΔGlu-155/Lys-208 variant could be fitted to a single phase or two-phase exponential regression with the same R^2 (0.9978) and residual sums of squares (0.0003) and a similar S.E. (0.0053 and 0.0058, respectively). However the two-phase exponential regression generated an unusually high k_1 of over 2200 min^{-1} and other unusual parameters, suggesting that a single phase exponential analysis is more appropriate for the ΔGlu-155/Lys-208 variant.

Therefore, the refolding curve of wild type and A140D enzymes monitored by fluorescence was better fitted by a two-phase exponential equation, which respectively gave 3.18 ± 0.57 and 3.34 ± 0.52 min^{-1} for k_1 and 0.22 ± 0.06 $min^{-1}$and 0.27 ± 0.05 min^{-1} for k_2, whereas the refolding of ΔGlu-155/Lys-208 variant followed single-phase exponential kinetics with a rate of 4.04 ± 0.80 min^{-1}. This suggests that the ΔGlu-155 variants still have a similar fast-refolding phase like the wild type and A140D enzymes but have lost the second slow refolding phase as shown in Fig. 6.

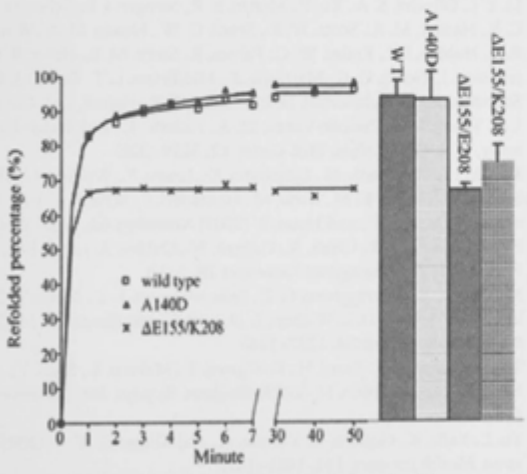


FIGURE 6. **Refolding of GSTO1 variants.** *Symbols* and *solid curves* represent the recovery of fluorescence. The *right-hand columns* show the recovery of activity with 4-nitrophenacylglutathione as a substrate (mean ± S.D.).

DISCUSSION

Previous studies have indicated that the polymorphic deletion of Glu-155 causes a deficiency of GSTO1-1 in cell lines, but the mechanism causing this deficiency is not clear. As the deletion can potentially alter splicing at the end of exon 4, we previously undertook Northern blotting analysis to eliminate a splicing defect as a cause of the GSTO1-1 deficiency (20). That study showed that T47D breast cancer cells that are hemizygous for the Glu-155 deletion allele produce mRNA of the appropriate size, suggesting that abnormal splicing is unlikely to be the cause of the deficiency. Previous pulse-chase studies in T47D cells indicate that the Glu-155 deletion protein has a shorter than normal $t_{1/2}$ (20). Although this variant can be expressed as an active enzyme in *E. coli*, the recombinant enzyme is relatively heat-sensitive (20). Given the clinical interest in GSTO1-1, we were interested in determining the mechanism by which the Glu-155 deletion mediates this deficiency.

The crystal structure indicated that the deletion of Glu-155 caused only small perturbations of the overall protein fold. Although there is a slight decrease in the volume and apparent accessibility of the H site, the distinct structural forms of ΔGlu-155 dimers observed in the crystal structures point toward increased flexibility of the protein. This may explain the previously observed increase in catalytic activity of ΔGlu-155 mutants (Table 4), as conformations suitable for binding and

TABLE 4

The activity of recombinant wild type and ΔGlu-155 GSTO1-1 variants with a range of substrates

All activities are shown as μmol/min/mg, mean ± S.D. CDNB, 1-chloro-2,4,dinitrobenzene; 2,4-DCPG; *S*-(2,4-dichlorophenacylglutathione). All data were compiled from Refs. 2, 8, and 19.

Enzyme	CDNB	Thiol transferase	Dehydroascorbate	Monomethyl arsonate	Dimethyl arsonate	2,4-DCPG
Human GSTO1-1	0.047 ± 0.01 0.039 ± 0.01	2 ± 0.21 2.6 ± 0.26	0.13 ± 0.005	0.33 ± 0.037	0.12 ± 0.006	1.3 ± 0.1
ΔGlu-155, Glu-208		6 ± 0.09	0.25 ± 0.045	0.65 ± 0.007	0.3 ± 0.014	1.3 ± 0.05
ΔGlu-155, Lys-208	0.063 ± 0.01	4.8 ± 0.07 3.6 ± 0.36	0.21 ± 0.03	0.67 ± 0.066	0.37 ± 0.015	1.1 ± 0.03

catalysis of diverse substrates are more accessible to the flexible mutant protein. The catalytic activity of the recombinant ΔGlu-155 enzymes and the absence of immunologically detectable GSTO1-1 in T47D cells suggest that the deletion may cause a defect in stability or folding rather than a loss of function. We found several structural factors that could have a deleterious impact on stability including the increase in the water accessible surface area, a reduction in buried hydrophobic surface area, and the large decrease in molecular contacts between helices α5 and α7. Even the increased buried area of dimer interface was not a positive factor because of the concomitant increase in the surface area of solvent-exposed hydrophobic residues.

The lower free energy change $\Delta G^{0}_{N\rightarrow U}$ supports the lower stability of the ΔGlu-155 enzymes. We also found evidence that at low urea concentrations the ΔGlu-155 enzymes form a transitional intermediate that is not significantly populated by the Glu-155-containing enzymes. Because their first transition is not dependent on the protein concentration, the unfolding intermediate formed by the ΔGlu-155 enzymes appears to be a dimer with more hydrophobic area exposure revealed by the ANS binding data. The nature of the intermediate formed by the Glu-155-containing enzymes is difficult to identify with the present data.

The capacity of the Glu-155 deletion variants to refold after complete unfolding in urea was limited compared with the Glu-155 containing Ala-140 and Asp-140 enzymes (Fig. 6). This limitation only occurred during the second slow phase of refolding. Because refolding is usually thought to be the reverse of unfolding, the refolding of the Glu-155 deletion enzymes is probably limited by the population of the transitional intermediate. As mentioned above, the intermediate has a larger exposed hydrophobic surface, which probably promotes precipitation or prevents further refolding of the protein resulting in the lack of the second slow phase for the ΔGlu-155 enzymes. Intracellular proteins are dynamic and may be subjected to conformational changes under different intracellular conditions. It is possible that the increased accessibility to solvent and decreased contacts in the Glu-155 deletion enzyme may promote unfolding and restrict refolding by presenting a barrier of increased hydrophobic surface exposure, leading to a dynamic equilibrium that favors an unfolded state *in vivo*.

We, therefore, conclude that the deletion of Glu-155 causes GSTO1-1 deficiency by its low stability, resulting from its increased propensity to unfold, and a novel mechanism that inhibits refolding. The structural characteristics of the deletion mutant, including a high accessible surface area and less buried hydrophobic area as well as altered packing around ΔGlu-155 agree with these observations. The results help us to understand the impact of the Glu-155 deletion allele on the level of GSTO1-1 activity *in vivo*, which may contribute to the observed association of GSTO1-1 with the age at onset of Alzheimer disease and the activation of the pro-inflammatory cytokine IL-1β.

Acknowledgments—We express our gratitude to Professor Mitchell Guss of the School of Molecular and Microbial Biosciences, University of Sydney, for the use of equipment during x-ray data collection.

REFERENCES

1. Board, P. G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermiin, L. S., Schulte, G. K., Danley, D. E., Hoth, L. R., Griffor, M. C., Kamath, A. V., Rosner, M. H., Chrunyk, B. A., Perregaux, D. E., Gabel, C. A., Geoghegan, K. F., and Pandit, J. (2000) *J. Biol. Chem.* **275,** 24798–24806
2. Whitbread, A. K., Tetlow, N., Eyre, H. J., Sutherland, G. R., and Board, P. G. (2003) *Pharmacogenetics* **13,** 131–144
3. Whitbread, A. K., Masoumi, A., Tetlow, N., Schmuck, E., Coggan, M., and Board, P. G. (2005) *Methods Enzymol.* **401,** 78–99
4. Zakharyan, R. A., Sampayo-Reyes, A., Healy, S. M., Tsaprailis, G., Board, P. G., Liebler, D. C., and Aposhian, H. V. (2001) *Chem. Res. Toxicol.* **14,** 1051–1057
5. Li, Y. J., Scott, W. K., Hedges, D. J., Zhang, F., Gaskell, P. C., Nance, M. A., Watts, R. L., Hubble, J. P., Koller, W. C., Pahwa, R., Stern, M. B., Hiner, B. C., Jankovic, J., Allen, F. A., Jr., Goetz, C. G., Mastaglia, F., Stajich, J. M., Gibson, R. A., Middleton, L. T., Saunders, A. M., Scott, B. L., Small, G. W., Nicodemus, K. K., Reed, A. D., Schmechel, D. E., Welsh-Bohmer, K. A., Conneally, P. M., Roses, A. D., Gilbert, J. R., Vance, J. M., Haines, J. L., and Pericak-Vance, M. A. (2002) *Am. J. Hum. Genet.* **70,** 985–993
6. Li, Y. J., Scott, W. K., Zhang, L., Lin, P. I., Oliveira, S. A., Skelly, T., Doraiswamy, M. P., Welsh-Bohmer, K. A., Martin, E. R., Haines, J. L., Pericak-Vance, M. A., and Vance, J. M. (2006) *Neurobiol. Aging* **27,** 1087–1093
7. Kölsch, H., Larionov, S., Dedeck, O., Orantes, M., Birkenmeier, G., Griffin, W. S., and Thal, D. R. (2007) *Stroke* **38,** 2847–2850
8. Board, P. G., and Anders, M. W. (2007) *Chem. Res. Toxicol.* **20,** 149–154
9. Board, P. G., Coggan, M., Cappello, J., Zhou, H., Oakley, A. J., and Anders, M. W. (2008) *Anal. Biochem.* **374,** 25–30
10. Laliberte, R. E., Perregaux, D. G., Hoth, L. R., Rosner, P. J., Jordan, C. K., Peese, K. M., Eggler, J. F., Dombroski, M. A., Geoghegan, K. F., and Gabel, C. A. (2003) *J. Biol. Chem.* **278,** 16567–16578
11. Dulhunty, A., Gage, P., Curtis, S., Chelvanayagam, G., and Board, P. (2001) *J. Biol. Chem.* **276,** 3319–3323
12. Li, Y. J., Oliveira, S. A., Xu, P., Martin, E. R., Stenger, J. E., Scherzer, C. R., Hauser, M. A., Scott, W. K., Small, G. W., Nance, M. A., Watts, R. L., Hubble, J. P., Koller, W. C., Pahwa, R., Stern, M. B., Hiner, B. C., Jankovic, J., Goetz, C. G., Mastaglia, F., Middleton, L. T., Roses, A. D., Saunders, A. M., Schmechel, D. E., Gullans, S. R., Haines, J. L., Gilbert, J. R., Vance, J. M., Pericak-Vance, M. A., Hulette, C., and Welsh-Bohmer, K. A. (2003) *Hum. Mol. Genet.* **12,** 3259–3267
13. Kölsch, H., Linnebank, M., Lütjohann, D., Jessen, F., Wüllner, U., Harbrecht, U., Thelen, K. M., Kreis, M., Hentschel, F., Schulz, A., von Bergmann, K., Maier, W., and Heun, R. (2004) *Neurology* **63,** 2255–2260
14. Paiva, L., Marcos, R., Creus, A., Coggan, M., Oakley, A. J., and Board, P. G. (2008) *Pharmacogenet. Genomics* **18,** 1–10
15. Mukherjee, B., Salavaggione, O. E., Pelleymounter, L. L., Moon, I., Eckloff, B. W., Schaid, D. J., Wieben, E. D., and Weinshilboum, R. M. (2006) *Drug. Metab. Dispos.* **34,** 1237–1246
16. Tanaka-Kagawa, T., Jinno, H., Hasegawa, T., Makino, Y., Seko, Y., Hanioka, N., and Ando, M. (2003) *Biochem. Biophys. Res. Commun.* **301,** 516–520
17. Yu, L., Kalla, K., Guthrie, E., Vidrine, A., and Klimecki, W. T. (2003) *Environ. Health Perspect.* **111,** 1421–1427
18. Marnell, L. L., Garcia-Vargas, G. G., Chowdhury, U. K., Zakharyan, R. A., Walsh, B., Avram, M. D., Kopplin, M. J., Cebrián, M. E., Silbergeld, E. K., and Aposhian, H. V. (2003) *Chem. Res. Toxicol.* **16,** 1507–1513
19. Schmuck, E. M., Board, P. G., Whitbread, A. K., Tetlow, N., Cavanaugh, J. A., Blackburn, A. C., and Masoumi, A. (2005) *Pharmacogenet. Genomics* **15,** 493–501
20. Schmuck, E., Cappello, J., Coggan, M., Brew, J., Cavanaugh, J. A., Blackburn, A. C., Baker, R. T., Eyre, H. J., Sutherland, G. R., and Board, P. G. (2008) *Int. J. Biochem. Cell Biol.* **40,** 2553–2559
21. Baker, R. T., Catanzariti, A. M., Karunasekara, Y., Soboleva, T. A., Sharwood, R., Whitney, S., and Board, P. G. (2005) *Methods Enzymol.* **398,** 540–554
22. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326

23. Vagin, A. A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
24. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119
25. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
26. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
27. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1948–1954
28. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucleic Acids Res.* **35,** W375–W383
29. Erhardt, J., and Dirr, H. (1995) *Eur. J. Biochem.* **230,** 614–620
30. Lakowicz, J. (1986) *Principles of Fluorescence Spectroscopy*, Plenum Press, New York
31. Szpikowska, B. K., and Mas, M. T. (1996) *Arch. Biochem. Biophys.* **335,** 173–182
32. Gloss, L. M., and Matthews, C. R. (1997) *Biochemistry* **36,** 5612–5623
33. Neet, K. E., and Timm, D. E. (1994) *Protein Sci.* **3,** 2167–2174
34. Grimsley, J. K., Scholtz, J. M., Pace, C. N., and Wild, J. R. (1997) *Biochemistry* **36,** 14366–14374
35. Krissinel, E., and Henrick, K. (2007) *J. Mol. Biol.* **372,** 774–797